DOCUMENT RESUME

ED 071 897 SE 015 534

TITLE	Project Physics Reader 4, Light and Electromagnetism.
INSTITUTION	Harvard Univ., Cambridge, Mass. Harvard Project Physics.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-5-1038
PUB DATE	67
CONTRACT	OEC-5-10-058
NOTE	254p.; Preliminary Version
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	Electricity; *Instructional Materials; Light; Magnets; *Physics; Radiation; *Science Materials; Secondary Grades; *Secondary School Science; *Supplementary Reading Materials
IDENTIFIERS	Harvard Project Physics

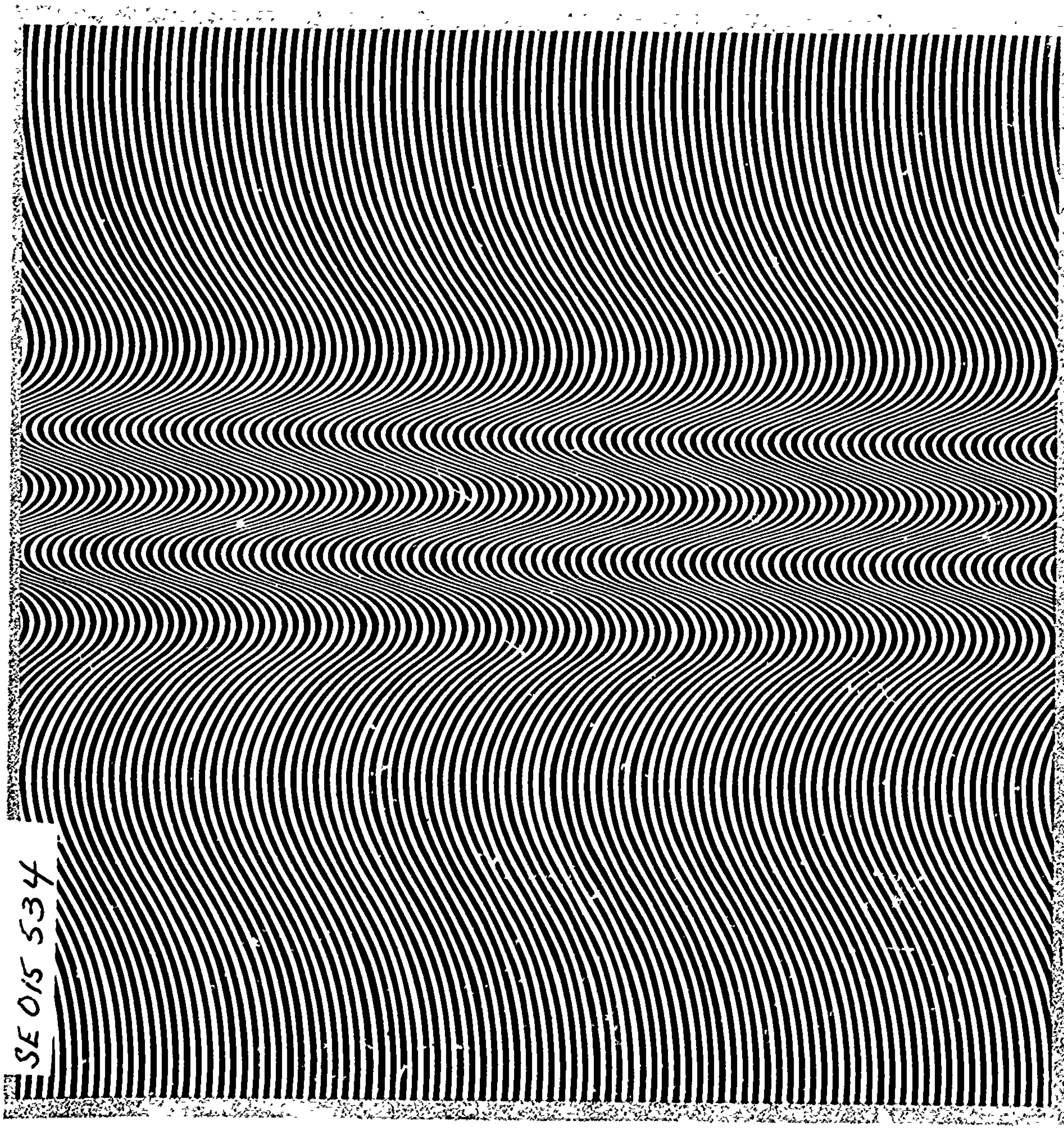
ABSTRACT

As a supplement to Project Physics Unit 4, a collection of articles is presented in this reader for student browsing. The 21 articles are included under the following headings: Letter from Thomas Jefferson; On the Method of Theoretical Physics; Systems, Feedback, Cybernetics; Velocity of Light; Popular Applications of Polarized Light; Eye and Camera; The laser--What it is and Does; A Simple Electric Circuit: Ohm's Law; The Electronic Revolution; The Invention of the Electric Light; High Fidelity; The Future of Direct Current Power Transmission; James Clerk Maxwell, Part II; On the Induction of Electric Currents; The Relationship of Electricity and Magnetism; The Electromagnetic Field; Radiation Belts Around the Earth; A Mirror for the Brain; Scientific Imagination; Lenses and Optical Instruments; and "Baffled!." Illustrations for explanation use are included. The work of Harvard Project Physics has been financially supported by: the Carnegie Corporation of New York, the Ford Foundation, the National Science Foundation, the Alfred P. Sloan Foundation, the United States Office of Education, and Harvard University. (CC)

Project Physics **Reader**

An Introduction to Physics

Light and Electromagnetism



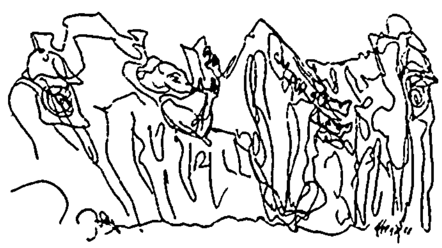
Project Physics **Reader**

An Introduction to Physics **4** Light and Electromagnetism

Preliminary Version **1967-68**

This is not a physics textbook. Rather, it is a physics reader, a collection of some of the best articles and book passages on physics. A few are on historic events in science, others contain some particularly memorable description of what physicists do; still others deal with philosophy of science, or with the impact of scientific thought on the imagination of the artist.

There are old and new classics, and also some little-known publications; many have been suggested for inclusion because some teacher or physicist remembered an article with particular fondness. The majority of articles is not drawn from scientific papers of historic importance themselves, because material from many of these is readily available, either as quotations in the Project Physics text or in special collections.



This collection is meant for your browsing. If you follow your own reading interests, chances are good that you will find here many pages that convey the joy these authors have in their work and the excitement of their ideas. If you want to follow up on interesting excerpts, the source list at the end of the reader will guide you for further reading.

Letter from Thomas Jefferson

June 1799

Monticello June 18. 99.

DEAR SIR,

I have to acknolege the reciept of your favor of May 14. in which you mention that you have finished the 6. first books of Euclid, plane trigonometry, surveying and algebra and ask whether I think a further pursuit of that branch of science would be useful to you. There are some propositions in the latter books of Euclid, and some of Archimedes, which are useful, and I have no doubt you have been made acquainted with them. Trigonometry, so far as this, is most valuable to every man, there is scarcely a day in which he will not resort to it for some of the purposes of common life; the science of calculation also is indispensible as far as the extraction of the square and cube roots; Algebra as far as the quadratic equation and the use of logarithms are often of value in ordinary cases; but all beyond these is but a luxury; a delicious luxury indeed; but not to be indulged in by one who is to have a profession to follow for his subsistence. In this light I view the conic sections, curves of the higher orders, perhaps even spherical trigonometry, Algebraical operations beyond the 2d dimension, and fluxions. There are other branches of science however worth the attention of every man: Astronomy, botany, chemistry, natural philosophy, natural history, anatomy. Not indeed to be a proficient in them; but to possess their general principles and outlines, so as that we may be able to amuse and inform ourselves further in any of them as we proceed through life and have occasion for them. Some knowlege of them is necessary for our character as well as comfort. The general elements of astronomy and of natural philosophy are best acquired at an academy where we can have the benefit of the instruments and apparatus usually provided there: but the others may well be acquired from books alone as far as our purposes require. I have indulged myself in these observations to you, because the evidence cannot be unuseful to you of a person who has often had occasion to consider which of his acquisitions in science have been really useful to him in life, and which of them have been merely a matter of luxury.

I am among those who think well of the human character generally. I consider man as formed for society, and endowed by nature with those dispositions which fit him for society. I believe also, with Condorcet, as mentioned in your letter, that his mind is perfectible to a degree of which we cannot as yet form any conception. It is impossible for a man who takes a survey of what is already known, not to see what an immensity in every branch of science yet remains to be discovered, and that too of articles to which our faculties seem adequate. In geometry and calculation we know a great deal. Yet there are some desiderata. In anatomy great progress has been made; but much is still to be acquired. In natural history we possess knowlege; but we want a great deal. In chemistry we are not yet sure of the first elements. Our natural philosophy is in a very infantine state; perhaps for great advances in it, a further progress in chemistry is necessary. Surgery is well advanced; but prodigiously short of what may be. The state of medecine is worse than that of total ignorance. Could we divest ourselves of every thing we suppose we know in it, we should start from a higher ground and with fairer prospects. From Hippocrates to Brown we have had nothing but a succession of hypothetical systems each having it's day of vogue, like the fashions and fancies of caps and gowns, and yielding in turn to the next caprice. Yet the human frame, which is to be the subject of suffering and torture under these learned modes, does not change. We have a few medecines, as the bark, opium, mercury, which in a few well defined diseases are of unquestionable virtue: but the residuary list of the materia medica, long as it is, contains but the charlataneries of the art; and of the diseases of doubtful form, physicians have ever had a false knowlege, worse than ignorance. Yet surely the list of unequivocal diseases and remedies is capable of enlargement; and it is still more certain that in the other branches of science, great fields are yet to be explored to which our faculties are equal, and that to an extent of which we cannot fix the limits. I join you therefore in branding as cowardly the idea that the human mind is incapable of further advances. This is precisely the doctrine which the present despots of the earth are inculcating, and their friends here re-echoing; and applying especially to religion and politics; 'that it is not probable that any thing better will be discovered than what was known to our fathers'. We are to look backwards then and not forwards for the improvement of science, and to find it amidst feudal barbarisms and the fires of Spital-fields. But thank heaven the American mind is already too much opened, to listen to these impostures; and while the art of printing is left to use, science can never be retrograde; what is once acquired of real knowlege can never be lost. To preserve the freedom of the human mind then and freedom of the press, every spirit should be ready to devote itself to martyrdom; for as long as we may think as we will, and speak as we think, the condition of man will proceed in improvement. The generation which is going off the stage has deserved well of mankind for the struggles it has made, and for having arrested that course of despotism which had overwhelmed the world for thousands and thousands of years. If there seems to be danger that the ground they have gained will be lost again, that danger comes from the generation your contemporary. But that the enthusiasm which characterises youth should lift it's parracide hands against freedom and science would be such a monstrous phaenomenon as I cannot place among possible things in this age and this country. Your college at least has shewn itself incapable of it; and if the youth of any other place have seemed to rally under other banners it has been from delusions which they will soon dissipate. I shall be happy to hear from you from time to time, and of your progress in study, and to be useful to you in whatever is in my power; being with sincere esteem Dear Sir

Your friend & servt
Th: Jefferson

On the Method of Theoretical Physics

Albert Einstein

An essay—1934.

IF YOU want to find out anything from the theoretical physicists about the methods they use, I advise you to stick closely to one principle: don't listen to their words, fix your attention on their deeds. To him who is a discoverer in this field the products of his imagination appear so necessary and natural that he regards them, and would like to have them regarded by others, not as creations of thought but as given realities.

These words sound like an invitation to you to walk out of this lecture. You will say to yourselves, the fellow's a working physicist himself and ought therefore to leave all questions of the structure of theoretical science to the epistemologists.

Against such criticism I can defend myself from the personal point of view by assuring you that it is not at my own instance but at the kind invitation of others that I have mounted this rostrum, which serves to commemorate a man who fought hard all his life for the unity of knowledge. Objectively, however, my enterprise can be justified on the ground that it may, after all, be of interest to know how one who has spent a life-time in striving with all his might to clear up and rectify its fundamentals looks upon his own branch of science. The way in which he regards its past and present may depend too much on what he hopes for the future and aims at in the present; but that is the inevitable fate of anybody who has occupied himself intensively with a world of ideas. The same thing happens to him as to the historian, who in the same way, even though perhaps unconsciously, groups actual events around ideals which he has formed for himself on the subject of human society.

Let us now cast an eye over the development of the theoretical system, paying special attention to the relations between the content of the theory and the totality of empirical fact. We are concerned with the eternal antithesis between the two inseparable components of our knowledge, the empirical and the rational, in our department.

We reverence ancient Greece as the cradle of western science. Here for the first time the world witnessed the miracle of a logical system which proceeded from step to step with such precision that every single one of its propositions was absolutely indubitable— I refer to Euclid's geometry. This admirable triumph of reasoning gave the human intellect the necessary confidence in itself for its subsequent achievements. If Euclid failed to kindle your youthful enthusiasm, then you were not born to be a scientific thinker.

But before mankind could be ripe for a science which takes in the whole of reality, a second fundamental truth was needed, which only became common property among philosophers with the advent of Kepler and Galileo. Pure logical thinking cannot yield us any knowledge of the empirical world; all knowledge of reality starts from experience and ends in it. Propositions arrived at by purely logical means are completely empty as regards reality. Because Galileo saw this, and particularly because he drummed it into the scientific world, he is the father of modern physics—indeed, of modern science altogether.

If, then, experience is the alpha and the omega of all our knowledge of reality, what is the function of pure reason in science?

A complete system of theoretical physics is made up of concepts, fundamental laws which are supposed to be valid for those concepts and conclusions to be reached by logical deduction. It is these conclusions which must correspond with our separate experiences; in any theoretical treatise their logical deduction occupies almost the whole book.

This is exactly what happens in Euclid's geometry, except that there the fundamental laws are called axioms and there is no question of the conclusions having to correspond to any sort of experience. If, however, one regard Euclidean geometry as the science of the possible mutual relations of practically rigid bodies in space, that is to say, treats it as a physical science, without abstracting from its original empirical content, the logical homogeneity of geometry and theoretical physics becomes complete.

We have thus assigned to pure reason and experience their places in a theoretical system of physics. The structure of the system is the work of reason; the empirical contents and their mutual relations must find their representation in the conclusions of the theory. In the possibility of such a representation lie the sole value and justification of the whole system, and especially of the concepts and fundamental principles which underlie it. These latter, by the way, are free inventions of the human intellect, which cannot be justified either by the nature of that intellect or in any other fashion *a priori.*

These fundamental concepts and postulates, which cannot be further reduced logically, form the essential part of a theory, which reason cannot touch. It is the grand object of all theory to make these irreducible elements as simple and as few in number as possible, without having to renounce the adequate representation of any empirical content whatever.

The view I have just outlined of the purely fictitious character of the fundamentals of scientific theory was by no means the prevailing one in the eighteenth or even the nineteenth century. But it is steadily gaining ground from the fact that the distance in thought between the fundamental concepts and laws on one side and, on the other, the conclusions which have to be brought into relation with our experience grows larger and larger, the simpler the logical structure becomes—that is to say, the smaller the number of logically independent conceptual elements which are found necessary to support the structure.

Newton, the first creator of a comprehensive, workable system of theoretical physics, still believed that the basic concepts and laws of his system could be derived from experience. This is no doubt the meaning of his saying, *hypotheses non fingo.*

Actually the concepts of time and space appeared at that time to present no difficulties. The concepts of mass, inertia and force, and the laws connecting them seemed to be drawn directly from experience. Once this basis is accepted, the expression for the force of gravitation appears derivable from experience, and it was reasonable to hope for the same in regard to other forces.

We can indeed see from Newton's formulation of it that the concept of absolute space, which comprised that of absolute rest, made him feel uncomfortable; he realized that there seemed to be nothing in experience corresponding to this last concept. He was also not quite comfortable about the introduction of forces operating at a distance. But the tremendous practical success of his doctrines may well have prevented him and the physicists of the eighteenth and nineteenth centuries from recognizing the fictitious character of the foundations of his system.

The natural philosophers of those days were, on the contrary, most of them possessed with the idea that the fundamental concepts and postulates of physics were not in the logical sense free inventions of the human mind but could be deduced from experience by "abstraction"—that is to say by logical means. A clear recognition of the erroneousness of this notion really only came with the general theory of relativity, which showed that one could take account of a wider range of empirical facts, and that too in a more satisfactory and complete manner, on a foundation quite different from the Newtonian. But quite apart from the question of the superiority of one or the other, the fictitious character of fundamental principles is perfectly evident from the fact that we can point to two essentially different principles, both of which correspond with experience to a large extent; this proves at the same time that every attempt at a logical deduction of the basic concepts and postulates of mechanics from elementary experiences is doomed to failure.

If, then, it is true that this axiomatic basis of theoretical physics cannot be extracted from experience but must be freely invented, can we ever hope to find the right way? Nay more, has this right way any existence outside our illusions? Can we hope to be guided in the right way by experience when there exist theories (such as classical mechanics) which to a large extent do justice to experience, without getting to the root of the matter? I answer without hesitation that there is, in my opinion, a right way, and that we are capable of finding it. Our experience hitherto justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas. I am convinced that we can discover by means of purely mathematical constructions the concepts and the laws connecting them with each other, which furnish the key to the understanding of natural phenomena. Experience may suggest the appropriate mathematical concepts, but they most certainly cannot be deduced from it. Experience remains, of course, the sole criterion of the physical utility of a mathematical construction. But the creative principle resides in mathematics. In a certain sense, therefore, I hold it true that pure thought can grasp reality, as the ancients dreamed.

In order to justify this confidence, I am compelled to make use of a mathematical conception. The physical world is represented as a four-dimensional continuum. If I assume a Riemannian metric in it and ask what are the simplest laws which such a metric system can satisfy, I arrive at the relativist theory of gravitation in empty space. If in that space I assume a vector-field or an anti-symmetrical tensor-field which can be inferred from it, and ask what are the simplest laws which such a field can satisfy, I arrive at Clerk Maxwell's equations for empty space.

At this point we still lack a theory for those parts of space in which electrical density does not disappear. De Broglie conjectured the existence of a wave field. which served to explain certain quantum properties of matter. Dirac found in the spinors field-magnitudes of a new sort, whose simplest equations enable one to a large extent to deduce the properties of the electron. Subsequently I discovered, in conjunction with my colleague, that these spinors form a special case of a new sort of field, mathematically connected with the four-dimensional system, which we called "semivectors." The simplest equations to which such semivectors can be reduced furnish a key to the understanding of the existence of two sorts of elementary particles, of different ponderable mass and equal but opposite electrical charge. These semivectors are, after ordinary vectors, the simplest mathematical fields that are possible in a metrical continuum of four dimensions, and it looks as if they described, in an easy manner, certain essential properties of electrical particles.

The important point for us to observe is that all these constructions and the laws connecting them can be arrived at by the principle of looking for the mathematically simplest concepts and the link between them. In the limited nature of the mathematically existent simple fields and the simple equations possible between them, lies the theorist's hope of grasping the real in all its depth.

Meanwhile the great stumbling-block for a field-theory of this kind lies in the conception of the atomic structure of matter and energy. For the theory is fundamentally non-atomic in so far as it operates exclusively with continuous functions of space, in contrast to classical mechanics, whose most important element, the material point, in itself does justice to the atomic structure of matter.

The modern quantum theory in the form associated with the names of de Broglie, Schrödinger, and Dirac, which operates with continuous functions, has overcome these difficulties by a bold piece of interpretation which was first given a clear form by Max Born. According to this, the spatial functions which appear in the equations make no claim to be a mathematical model of the atomic structure. Those functions are only supposed to determine the mathematical probabilities of the occurrence of such structures if measurements were taken at a particular spot or in a certain state of motion. This notion is logically unobjectionable and has important successes to its credit. Unfortunately, however, it compels one to use a continuum the number of whose dimensions is not that ascribed to space by physics hitherto (four) but rises indefinitely with the number of the particles constituting the system under consideration. I cannot but confess that I attach only a transitory importance to this interpretation. I still believe in the possibility of a model of reality—that is to say, of a theory which represents things themselves and not merely the probability of their occurrence.

On the other hand it seems to me certain that we must give up the idea of a complete localization of the particles in a theoretical model. This seems to me to be the permanent upshot of Heisenberg's principle of uncertainty. But an atomic theory in the true sense of the word (not merely on the basis of an interpretation) without localization of particles in a mathematical model, is perfectly thinkable. For instance, to account for the atomic character of electricity, the field equations need only lead to the following conclusions: A portion of space (three-dimensional) at whose boundaries electrical density disappears everywhere, always contains a total electrical charge whose size is represented by a whole number. In a continuum-theory atomic characteristics would be satisfactorily expressed by integral laws without localization of the formation entity which constitutes the atomic structure.

Not until the atomic structure has been successfully represented in such a manner would I consider the quantum-riddle solved.

Experiments and Calculations Relative to Physical Optics

Thomas Young

A scientific paper published in 1855.

EXPERIMENTS AND CALCULATIONS RELATIVE TO

PHYSICAL OPTICS.

From the Philosophical Transactions for 1804.

A BAKERIAN LECTURE.

READ Nov. 24, 1803.

I.—*Experimental Demonstration of the general Law of the Interference of Light.*

IN making some experiments on the fringes of colours accompanying shadows, I have found so simple and so demonstrative a proof of the general law of the interference of two portions of light, which I have already endeavoured to establish, that I think it right to lay before the Royal Society a short statement of the facts which appear to me so decisive. The proposition on which I mean to insist, at present, is simply this—that fringes of colours are produced by the interference of two portions of light; and I think it will not be denied by the most prejudiced, that the assertion is proved by the experiments I am about to relate, which may be repeated with great ease whenever the sun shines, and without any other apparatus than is at hand to every one.

Exper. 1. I made a small hole in a window-shutter, and covered it with a piece of thick paper, which I perforated with a fine needle. For greater convenience of observation I placed a small looking-glass without the window-shutter, in such a position as to reflect the sun's light, in a direction nearly horizontal, upon the opposite wall, and to cause the cone of diverging light to pass over a table on which were several little screens of card-paper. I brought into the sunbeam a slip of card, about one-thirtieth of an inch in breadth, and observed its shadow, either on the wall or on other cards held at different distances. Besides the fringes of colour on each side of the shadow, the shadow itself was divided by similar parallel fringes, of smaller dimensions, differing in number, according to the distance at which the shadow was observed. but leaving the middle of the shadow always white. Now these fringes were the joint effects of the portions of light passing on each side of the slip of card, and inflected, or rather diffracted, into the shadow. For, a little screen being placed a few inches from the card, so as to receive either edge of the shadow on its margin, all the fringes which had before been observed in the shadow on the wall, immediately disappeared, although the light inflected on the other side was allowed to retain its course, and although this light must have undergone any modification that the proximity of the other edge of the slip of card might have been capable of occasioning. When the interposed screen was more remote from the narrow card, it was necessary to plunge it more deeply into the shadow, in order to extinguish the parallel lines; for here the light, diffracted from the edge of the object, had entered further into the shadow in its way towards the fringes. Nor was it for want of a sufficient intensity of light that one of the two portions was incapable of producing the fringes alone; for, when they were both uninterrupted, the lines appeared, even if the intensity was reduced to one-tenth or one-twentieth.

Exper. 2. The crested fringes described by the ingenious and accurate Grimaldi, afford an elegant variation of the preceding experiment, and an interesting example of a calculation grounded on it. When a shadow is formed by an object which has a rectangular termination, besides the usual external fringes, there are two or three alternations of colours. beginning from the line which bisects the angle, disposed on each side of it in curves, which are convex towards the bisecting line, and which converge in some degree towards it, as they become more remote from the angular point. These fringes are also the joint effect of the light which is inflected directly towards the shadow from each of the two outlines of the object; for if a screen be placed within a few inches of the object, so as to receive only one of the edges of the shadow, the whole of the fringes disappear : if, on the contrary, the rectangular point of the screen be opposed to the point of the shadow, so as barely to receive the angle of the shadow on its extremity, the fringes will remain undisturbed.

II.— *Comparison of Measures deduced from various Experiments.*

If we now proceed to examine the dimensions of the fringes, under different circumstances, we may calculate the differences of the lengths of the paths described by the portions of light which have thus been proved to be concerned in producing those fringes ; and we shall find that, where the lengths are equal, the light always remains white; but that, where either the brightest light, or the light of any given colour, disappears and reappears, a first, a second, or a third time, the differences of the lengths of the paths of the two portions are in arithmetical progression, as nearly as we can expect experiments of this kind to agree with each other. I shall compare, in this point of view, the measures deduced from several experiments of Newton, and from some of my own.

In the eighth and ninth observations of the third book of Newton's Optics, some experiments are related, which, together with the third observation, will furnish us with the data necessary for the calculation. Two knives were placed, with their edges meeting at a very acute angle, in a beam of the sun's light, admitted through a small aperture, and the point of concourse of the two first dark lines bordering the shadows of the respective knives was observed at various distances. The results of six observations are expressed in the first three lines of the first Table. On the supposition that the dark line is produced by the first interference of the light reflected from the edges of the knives, with the light passing in a straight line between them, we may assign, by calculating the difference of the two paths, the interval for the first disappearance of the brightest light, as it is expressed in the fourth line. The second Table contains the results of a similar calculation, from Newton's observations on the shadow of a hair ; and the third, from some experiments of my own, of the same nature; the second bright line being supposed to correspond to a double interval, the second dark line to a triple interval, and the succeeding lines to depend on a continuation of the progression. The unit of all the tables is an inch.

TABLE I. *Obs.* 9. N.

Distance of the knives from the aperture						101
Distance of the paper from the knives .	$1\frac{1}{2}$,	$3\frac{1}{3}$,	$8\frac{1}{3}$,	32,	96.	131
Distances between the edges of the knives, opposite to the point of concourse .	.012,	.020,	.034,	.057,	.081,	.087
Interval of disappearance . .	.0000122,	.0000155,	.0000182,	.0000167,	.0000166,	.0000166

TABLE II. *Obs.* 3. N.

Breadth of the hair		$\frac{1}{280}$
Distance of the hair from the aperture . . .		144
Distances of the scale from the aperture . . .	150.	252
(Breadths of the shadow	$\frac{1}{30}$,	$\frac{1}{9}$)
Breadth between the second pair of bright lines .	$\frac{2}{47}$,	$\frac{4}{17}$
Interval of disappearance, or half the difference of the paths	.0000151,	.0000173
Breadth between the third pair of bright lines . .	$\frac{4}{73}$,	$\frac{3}{16}$
Interval of disappearance, one-fourth of the difference	.0000130,	.0000143

TABLE III. *Exper.* 3.

Breadth of the object	.434
Distance of the object from the aperture	125
Distance of the wall from the aperture	250
Distance of the second pair of dark lines from each other . .	1.167
Interval of disappearance, one-third of the difference . . .	.0000149

Exper. 4.

Breadth of the wire	.083
Distance of the wire from the aperture	32
Distance of the wall from the aperture	250
(Breadth of the shadow, by three measurements	.815, .826, or .827; mean, .823)
Distance of the first pair of dark lines	1.165, 1.170, or 1.160; mean, 1.165
Interval of disappearance . . .	.0000194
Distance of the second pair of dark lines	1.402, 1.395, or 1.400; mean, 1.399
Interval of disappearance . . .	.0000137
Distance of the third pair of dark lines	1.594, 1.580, or 1.585; mean, 1.586
Interval of disappearance . . .	.0000128

It appears, from five of the six observations of the first Table, in which the distance of the shadow was varied from about 3 inches to 11 feet, and the breadth of the fringes was increased in the ratio of 7 to 1, that the difference of the routes constituting the interval of disappearance, varied but one-eleventh at most; and that, in three out of the five, it agreed with the mean, either exactly, or within $\frac{1}{100}$ part. Hence we are warranted in inferring that the interval appropriate to the extinction of the brightest light, is either accurately or very nearly constant.

But it may be inferred, from a comparison of all the other observations, that when the obliquity of the reflection is very great, some circumstance takes place, which causes the interval thus calculated to be somewhat greater: thus, in the eleventh line of the third Table, it comes out one-sixth greater than the mean of the five already mentioned. On the other hand, the mean of two of Newton's experiments and one of mine, is a result about one-fourth less than the former. With respect to the nature of this circumstance, I cannot at present form a decided opinion; but I conjecture that it is a deviation of some of the light concerned, from the rectilinear direction assigned to it, arising either from its natural diffraction, by which the magnitude of the shadow is also enlarged, or from some other unknown cause. If we imagined the shadow of the wire, and the fringes nearest it, to be so contracted that the motion of the light bounding the shadow might be rectilinear, we should thus make a sufficient compensation for this deviation; but it is difficult to point out what precise track of the light would cause it to require this correction.

The mean of the three experiments which appear to have been least affected by this unknown deviation, gives .0000127 for the interval appropriate to the disappearance of the brightest light; and it may be inferred that if they had been wholly exempted from its effects, the measure would have been somewhat smaller. Now the analogous interval, deduced from the experiments of Newton on thin plates, is .0000112, which is about one-eighth less than the former result; and this appears to be a coincidence fully sufficient to authorise us to attribute these two classes of phenomena to the same cause. It is very easily shown, with respect to the colours of thin plates, that each kind of light disappears and reappears where the differences of the routes of two of its portions are in arithmetical progression; and we have seen that the same law may be in general inferred from the phenomena of diffracted light, even independently of the analogy.

The distribution of the colours is also so similar in both cases, as to point immediately to a similarity in the causes. In the thirteenth observation of the second part of the first book, Newton relates, that the interval of the glasses where the rings appeared in red light, was to the interval where they appeared in violet light, as 14 to 9; and, in the eleventh observation of the third book, that the distances between the fringes, under the same circumstances, were the 22d and 27th of an inch. Hence, deducting the breadth of the hair and taking the squares, in order to find the relation of the difference of the routes, we have the proportion of 14 to $9\frac{1}{4}$, which scarcely differs from the proportion observed in the colours of the thin plate.

We may readily determine, from this general principle, the form of the crested fringes of Grimaldi, already described; for it will appear that, under the circumstances of the experiment related, the points in which the differences of the lengths of the paths described by the two portions of light are equal to a constant quantity, and in which, therefore, the same kinds of light ought to appear or disappear, are always found in equilateral hyperbolas, of which the axes coincide with the outlines of the shadow, and the asymptotes nearly with the diagonal line. Such, therefore, must be the direction of the fringes; and this conclusion agrees perfectly with the observation. But it must be remarked, that the parts near the outlines of the shadow are so much shaded off, as to render the character of the curve somewhat less decidedly marked where it approaches to its axis. These fringes have a slight resemblance to the hyperbolic fringes observed by Newton; but the analogy is only distant.

III.—*Application to the Supernumerary Rainbows*.

The repetitions of colours sometimes observed within the common rainbow, and described in the Philosophical Transactions, by Dr. Langwith and Mr. Daval, admit also a very easy and complete explanation from the same principles, Dr. Pemberton has attempted to point out an analogy between these colours and those of thin plates; but the irregular reflection from the posterior surface of the drop, to which alone he attributes the appearance, must be far too weak to produce any visible effects. In order to understand the phenomenon, we have only to attend to the two portions of light which are exhibited in the common diagrams explanatory of the rainbow, regularly reflected from the posterior surface of the drop, and crossing each other in various directions, till, at the angle of the greatest deviation, they coincide with each other, so as to produce, by the greater intensity of this redoubled light, the common rainbow of 41 degrees. Other parts of these two portions will quit the drop in directions parallel to each other; and these would exhibit a continued diffusion of fainter light, for 25° within the bright termination which forms the rainbow, but for the general law of interference, which, as in other similar cases, divides the light into concentric rings; the magnitude of these rings depending on that of the drop, according to the difference of time occupied in the passage of the two portions, which thus proceed in parallel directions to the spectator's eye, after having been differently refracted and reflected within the drop. This difference varies, at first, nearly as the square of the angular distance from the primitive rainbow; and, if the first additional red be at the distance of 2° from the red of the rainbow, so as to interfere a little with the primitive violet, the fourth additional red will be at a distance of nearly 2° more; and the intermediate colours will occupy a space nearly equal to the original rainbow. In order to produce this effect, the drops must be about $\frac{1}{76}$ of an inch, or .013, in diameter: it would be sufficient if they were between $\frac{1}{70}$ and $\frac{1}{80}$. The reason that such supernumerary colours are not often seen, must be, that it does not often happen that drops so nearly equal are found together; but, that this may sometimes happen, is not in itself at all improbable: we measure even medicines by dropping them from a phial, and it may easily be conceived that the drops formed by natural operations may sometimes be as uniform as any that can be produced by art. How accurately this theory coincides with the observation, may best be determined from Dr. Langwith's own words.

"August the 21st, 1722, about half an hour past five in the evening, weather temperate, wind at north-east, the appearance was as follows:—The colours of the primary rainbow were as usual, only the purple very much inclining to red, and well defined: under this was an arch of green, the upper part of which inclined to a bright yellow, the lower to a more dusky green: under this were alternately two arches of reddish purple, and two of green: under all, a faint appearance of another arch of purple, which vanished and returned several times so quick, that we could not readily fix our eyes upon it. Thus the order of the colours was, I. Red, orange-colour, yellow, green, light-blue, deep blue, purple. II. Light green, dark green, purple. III. Green, purple. IV. Green, faint vanishing purple. You see we had here four orders of colours, and perhaps the beginning of a fifth: for I make no question but that what I call the purple, is a mixture of the purple of each of the upper series with the red of the next below it, and the green a mixture of the intermediate colours. I send you not this account barely upon the credit of my own eyes; for there was a clergyman and four other gentlemen in company, whom I desired to view the colours attentively, who all agreed that they appear in the manner that I have now described. There are two things which well deserve to be taken notice of, as they may perhaps direct us, in some measure, to the solution of this curious phenomenon The first is, that the breadth of the first series so far exceeded that of any of the rest, that, as near as I could judge, it was equal to them all taken together. The second is, that I have never observed these inner orders of colours in the lower parts of the rainbow, though they have often been incomparably more vivid than the upper parts, under which the colours have appeared. I have taken notice of this so very often, that I can hardly look upon it to be accidental;

and, if it should prove true in general, it will bring the disquisition into a narrow compass; for it will show that this effect depends upon some property which the drops retain, whilst they are in the upper part of the air, but lose as they come lower, and are more mixed with one another." Phil. Trans., Vol. XXXII. p. 243.

From a consideration of the nature of the secondary rainbow, of 54°, it may be inferred, that if any such supernumerary colours were seen attending this rainbow, they would necessarily be external to it, instead of internal. The circles sometimes seen encompassing the observer's shadow in a mist, are perhaps more nearly related to the common colours of thin plates as seen by reflection.

IV.—*Argumentative Inference respecting the Nature of Light.*

The experiment of Grimaldi, on the crested fringes within the shadow, together with several others of his observations, equally important, has been left unnoticed by Newton. Those who are attached to the Newtonian theory of light, or to the hypothesis of modern opticians, founded on views still less enlarged, would do well to endeavour to imagine any thing like an explanation of these experiments, derived from their own doctrines; and, if they fail in the attempt, to refrain at least from idle declamation against a system which is founded on the accuracy of its application to all these facts, and to a thousand others of a similar nature.

From the experiments and calculations which have been premised, we may be allowed to infer, that homogeneous light, at certain equal distances in the direction of its motion, is possessed of opposite qualities, capable of neutralising or destroying each other, and of extinguishing the light, where they happen to be united; that these qualities succeed each other alternately in successive concentric superficies, at distances which are constant for the same light, passing through the same medium. From the agreement of the measures, and from the similarity of the phenomena, we may conclude, that these intervals are the same as are concerned in the production of the colours of thin plates;

but these are shown, by the experiments of Newton, to be the smaller, the denser the medium; and, since it may be presumed that their number must necessarily remain unaltered in a given quantity of light, it follows of course, that light moves more slowly in a denser, than in a rarer medium; and this being granted, it must be allowed, that refraction is not the effect of an attractive force directed to a denser medium. The advocates for the projectile hypothesis of light, must consider which link in this chain of reasoning they may judge to be the most feeble; for, hitherto, I have advanced in this Paper no general hypothesis whatever. But, since we know that sound diverges in concentric superficies, and that musical sounds consist of opposite qualities, capable of neutralising each other, and succeeding at certain equal intervals, which are different according to the difference of the note, we are fully authorized to conclude, that there must be some strong resemblance between the nature of sound and that of light.

I have not, in the course of these investigations, found any reason to suppose the presence of such an inflecting medium in the neighbourhood of dense substances as I was formerly inclined to attribute to them; and, upon considering the phenomena of the aberration of the stars, I am disposed to believe that the luminiferous ether pervades the substance of all material bodies with little or no resistance, as freely perhaps as the wind passes through a grove of trees.

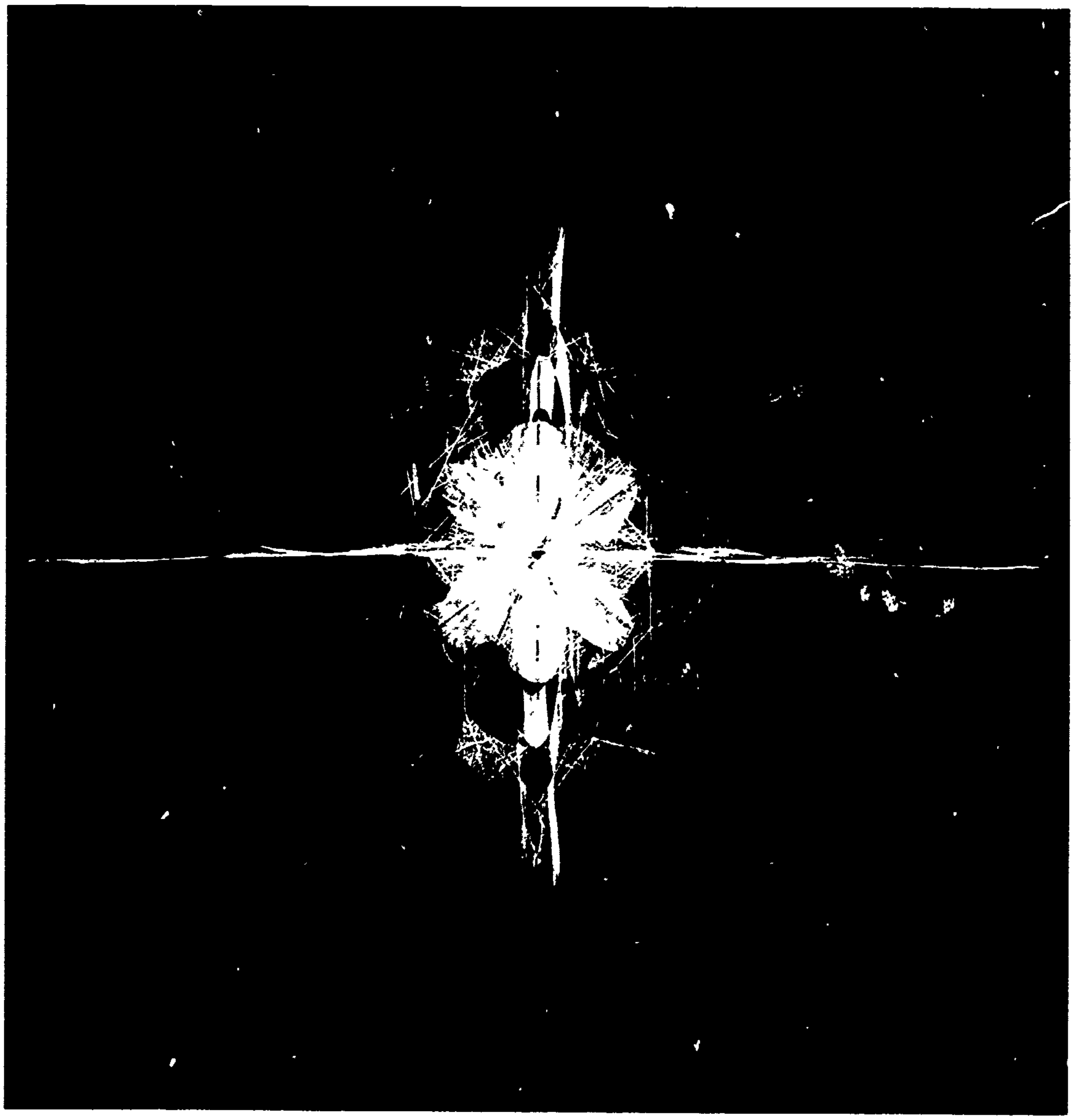
The observations on the effects of diffraction and interference may perhaps sometimes be applied to a practical purpose, in making us cautious in our conclusions respecting the appearances of minute bodies viewed in a microscope. The shadow of a fibre, however opaque, placed in a pencil of light admitted through a small aperture, is always somewhat less dark in the middle of its breadth than in the parts on each side. A similar effect may also take place, in some degree, with respect to the image on the retina, and impress the sense with an idea of a transparency which has no real existence: and if a small portion of light be really transmitted through the substance, this may again be destroyed by its interference with the diffracted light, and produce an appearance of partial opacity, instead of uniform semi-transparency. Thus a central dark spot and a light spot, surrounded by a darker circle, may respectively be produced in the images of a semi-transparent and an opaque corpuscle: and impress us with an idea of a complication of structure which does not exist. In order to detect the fallacy, we may make two or three fibres cross each other, and view a number of globules contiguous to each other; or we may obtain a still more effectual remedy by changing the magnifying power; and then, if the appearance remain constant in kind and in degree, we may be assured that it truly represents the nature of the substance to be examined. It is natural to inquire whether or no the figures of the globules of blood, delineated by Mr. Hewson in the Phil. Trans., Vol. LXIII. for 1773, might not in some measure have been influenced by a deception of this kind: but, as far as I have hitherto been able to examine the globules, with a lens of one-fiftieth of an inch focus, I have found them nearly such as Mr. Hewson has described them.

V.—*Remarks on the Colours of Natural Bodies.*

Exper. 5. I have already adduced, in illustration of Newton's comparison of the colours of natural bodies with those of thin plates, Dr. Wollaston's observations on the blue light of the lower part of a candle, which appears, when viewed through a prism, to be divided into five portions. I have lately observed a similar instance, still more strongly marked, in the light transmitted by the blue glass sold by the opticians. This light is separated by the prism into seven distinct portions, nearly equal in magnitude, but somewhat broader, and less accurately defined, towards the violet end of the spectrum. The first two are red, the third is yellowish green, the fourth green, the fifth blue, the sixth bluish violet, and the seventh violet. This division agrees very nearly with that of the light reflected by a plate of air $\frac{1}{2340}$ of an inch in thickness, corresponding to the 11th series of red, and the 18th of violet. A similar plate of a metallic oxide would perhaps be about $\frac{1}{11600}$ of an inch in thickness. But it must be confessed that there are strong reasons for believing the colouring particles of natural bodies in general to be incomparably smaller than this; and it is probable that the analogy suggested by Newton is somewhat less close than he imagined. The light reflected by a plate of air, at any thickness nearly corresponding to the 11th red, appears to the eye to be very nearly white; but, under favourable circumstances, the 11th red and the neighbouring colours may still be distinguished. The light of some kinds of coloured glass is pure red; that of others red with a little green: some intercept all the light, except the extreme red and the blue. In the blue light of a candle, expanded by the prism, the portions of each colour appear to be narrower, and the intervening dark spaces wider than in the analogous spectrum derived from the light reflected from a thin plate. The light of burning alcohol appears to be green and violet only. The pink dye sold in the shops, which is a preparation of the carthamus, affords a good specimen of a yellow green light regularly reflected, and a crimson probably produced by transmission.

VI.—*Experiment on the Dark Rays of Ritter.*

Exper. 6. The existence of solar rays accompanying light, more refrangible than the violet rays, and cognisable by their chemical effects, was first ascertained by Mr. Ritter; but Dr. Wollaston made the same experiments a very short time afterwards, without having been informed of what had been done on the Continent. These rays appear to extend beyond the violet rays of the prismatic spectrum, through a space nearly equal to that which is occupied by the violet. In order to complete the comparison of their properties with those of visible light, I was desirous of examining the effect of their reflection from a thin plate of air, capable of producing the well-known rings of colours. For this purpose I formed an image of the rings, by means of the solar microscope, with the apparatus which I have described in the Journals of the Royal Institution, and I threw this image on paper dipped in a solution of nitrate of silver, placed at the distance of about nine inches from the microscope. In the course of an hour portions of three dark rings were very distinctly visible, much smaller than the brightest rings of the coloured image, and coinciding very nearly, in their dimensions, with the rings of violet light that appeared upon the interposition of violet glass. I thought the dark rings were a little smaller than the violet rings, but the difference was not sufficiently great to be accurately ascertained; it might be as much as $\frac{1}{30}$ or $\frac{1}{40}$ of the diameters, but not greater. It is the less surprising that the difference should be so small, as the dimensions of the coloured rings do not by any means vary at the violet end of the spectrum so rapidly as at the red end. For performing this experiment with very great accuracy a heliostate would be necessary, since the motion of the sun causes a slight change in the place of the image; and leather, impregnated with the muriate of silver, would indicate the effect with greater delicacy. The experiment, however, in its present state, is sufficient to complete the analogy of the invisible with the visible rays, and to show that they are equally liable to the general law which is the principal subject of this Paper. If we had thermometers sufficiently delicate, it is probable that we might acquire, by similar means, information still more interesting, with respect to the rays of invisible heat discovered by Dr. Herschel; but at present there is great reason to doubt of the practicability of such an experiment.



Variation within a Sphere, No. 10: The Sun.
Sculptural construction of gold wire, 22 feet long, 11 feet high, 5-1/2 feet deep. By Richard Lippold, American sculptor.
The Metropolitan Museum of Art, N.Y.C.

Velocity of Light

A. A. Michelson

A chapter from his book, *Studies in Optics,* published in 1927.

The velocity of light is one of the most important of the fundamental constants of Nature. Its measurement by Foucault and Fizeau gave as the result a speed greater in air than in water, thus deciding in favor of the undulatory and against the corpuscular theory. Again, the comparison of the electrostatic and the electromagnetic units gives as an experimental result a value remarkably close to the velocity of light—a result which justified Maxwell in concluding that light is the propagation of an electromagnetic disturbance. Finally, the principle of relativity gives the velocity of light a still greater importance, since one of its fundamental postulates is the constancy of this velocity under all possible conditions.

The first attempt at measurement was due to Galileo. Two observers, placed at a distance of several kilometers, are provided with lanterns which can be covered or uncovered by a movable screen. The first observer uncovers his light, and the second observer answers by uncovering his at the instant of perceiving the light from the first. If there is an interval between the uncovering of the lantern by the first observer and his perception of the return signal from the second (due allowance being made for the delay between perception and motion), the distance divided by the time interval should give the velocity of propagation.

Needless to say, the time interval was far too small to be appreciated by such imperfect appliances. It is nevertheless worthy of note that the principle of the method is sound, and, with improvements that are almost intuitive, leads to the well-known method of Fizeau. The first improvement would clearly be the substitution of a mirror instead of the second observer. The second would consist in the substitution of a series of equidistant apertures in a rapidly revolving screen instead of the single screen which covers and uncovers the light.

The first actual determination of the velocity of light was made in 1675 by Römer as a result of his observation of the eclipses of the first satellite of Jupiter. These eclipses, recurring at very nearly equal intervals, could be calculated, and Römer found that the observed and the calculated values showed an annual discrepancy. The eclipses were later by an interval of sixteen minutes and twenty-six seconds[1] when the earth is farthest from Jupiter than when nearest to it. Römer correctly attributed this difference to the time required by light to traverse the earth's orbit. If this be taken as 300,000,000 kilometers and the time interval as one thousand seconds, the resulting value for the velocity of light is 300,000 kilometers per second.

Another method for the determination of the velocity of light is due to Bradley, who in 1728 announced an apparent annual deviation in the direction of the fixed stars from their mean position, to which he gave the name "aberration." A star whose direction is at right angles to the earth's orbital motion appears displaced in the direc-

[1] The value originally given by Römer, twenty-two minutes, is clearly too great.

tion of motion by an angle of 20″.445. This displacement Bradley attributed to the finite velocity of light.

With a telescope pointing in the true direction of such a star, during the time of passage of the light from objective to focus the telescope will have been displaced in consequence of the orbital motion of the earth so that the image of the star falls behind the crosshairs. In order to produce coincidence, the telescope must be inclined forward at such an angle α that the tangent is equal to the ratio of the velocity v of the earth to the velocity of light,

$$\tan \alpha = \frac{v}{V},$$

or, since $v = \pi D/T$, where D is the diameter of the earth's orbit and T the number of seconds in the year,

$$\tan \alpha = \frac{\pi D}{VT},$$

from which the velocity of light may be found; but, as is also the case with the method of Römer, only to the degree of accuracy with which the sun's distance, $\frac{1}{2}D$, is known; that is, with an order of accuracy of about 1 per cent.[1]

In 1849 Fizeau announced the result of the first experimental measurement of the velocity of light. Two astronomical telescope objectives L_1 and L_2 (Fig. 73) are placed facing each other at the two stations. At the focus of the first is an intense but minute image a of the source of light (arc) by reflection from a plane-parallel plate N.

[1] The value of the velocity of light has been obtained, by experimental methods immediately to be described, with an order of accuracy of one in one hundred thousand, so that now the process is inverted, and this result is employed to find the sun's distance.

The light from this image is rendered approximately parallel by the first objective. These parallel rays, falling on the distant objective, are brought to a focus at the surface of a mirror, whence the path is retraced and an image formed which coincides with the original image *a*, where it is observed by the ocular *E*. An accurately

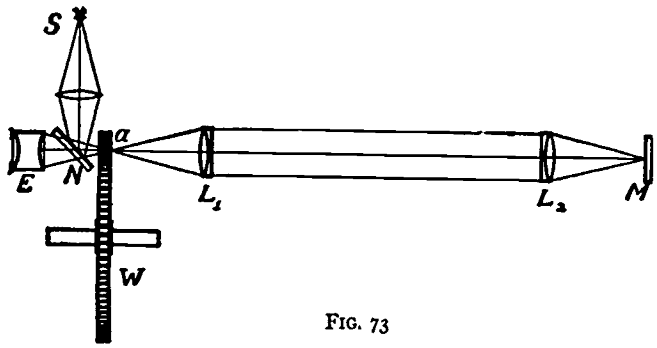


FIG. 73

divided toothed wheel *W* is given a uniform rotation, thus interrupting the passage of the light at *a*. If, on returning, the light is blocked by a tooth, it is eclipsed, to reappear at a velocity such that the next succeeding interval occupies the place of the former, and so on.

If n is the number of teeth and N the number of turns per second, K the number of teeth which pass during the double journey of the light over the distance D,

$$V=\frac{2NnD}{K}.$$

It is easier to mark the minima than the maxima of intensity, and accordingly

$$K=\frac{2p-1}{2}$$

if p is the order of the eclipse. Let δK be the error committed in the estimate of K (practically the error in estimation of equality of intensities on the descending and the ascending branches of the intensity curve). Then

$$\frac{\partial V}{V}=\frac{\partial K}{K}.$$

Hence it is desirable to make K as great as possible. In Fizeau's experiments this number was 5 to 7, and should have given a result correct to about one three-hundredth. It was, in fact, about 5 per cent too large.

A much more accurate determination was undertaken by Cornu in 1872 in which K varied from 3 to 21, the result as given by Cornu being 300,400, with a probable error of one-tenth of 1 per cent. In discussing Cornu's results, however, Listing showed that these tended toward a smaller value as the speed increased, and he assigns this limit as the correct value, namely, 299,950. Perrotin, with the same apparatus, found 299,900.

Before Fizeau had concluded his experiments, another project was proposed by Arago, namely, the utilization of the revolving mirror by means of which Wheatstone had measured the speed of propagation of an electric current. Arago's chief interest in the problem lay in the possibility of deciding the question of the relative velocities in air and water as a crucial test between the undulatory and the corpuscular theories. He pointed out, however, the possibility of measuring the absolute velocity.

The plan was to compare the deviations of the light from an electric spark reflected directly from the revolving mirror with that which was reflected after traversing a considerable distance in air (or in water). The difficulty in executing such an experiment lay in the uncertainty in the direction in which the two reflected images of the spark were to appear (which might be anywhere in 360°). This difficulty was solved by Foucault in 1862 by the following ingenious device whereby the return light is

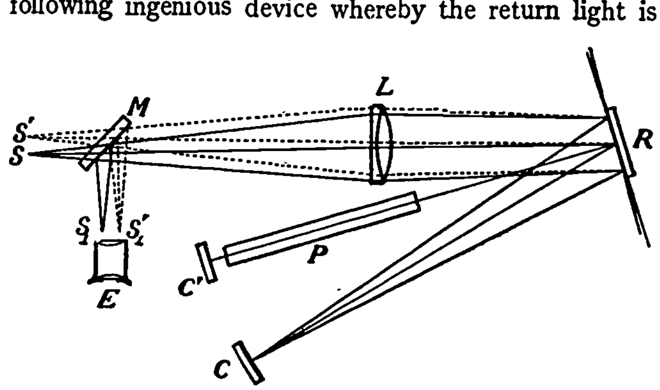


FIG. 74

always reflected in the same direction (apart from the deviation due to the retardation which it is required to measure), notwithstanding the rotation of the mirror.

Following is the actual arrangement of apparatus by which this is effected. Light from a source S falls upon an objective L, whence it proceeds to the revolving mirror R, and is thence reflected to the concave mirror C (whose center is at R), where it forms a real image of the source. It then retraces its path, forming a real image which coincides with the source even when the revolving mirror is in slow motion. Part of the light is reflected from the plane-parallel glass M, forming an image at a where it is observed by the micrometer eyepiece E.

If now the revolving mirror is turning rapidly, the return image, instead of coinciding with its original position, will be deviated in the direction of rotation through an angle double that through which the mirror turns while the light makes its double transit. If this angle is a and the distance between mirrors is D, and the revolving mirror makes N turns per second,

$$a = 2\pi N \frac{2D}{V},$$

or

$$V = \frac{4\pi N D}{a}.$$

In principle there is no essential difference between the two methods. In the method of the toothed wheel the angle a corresponds to the passage of K teeth, and is therefore $a = 2\pi K/n$, so that the formula previously found, $V = \frac{2NnD}{K}$, now becomes $V = \frac{4\pi ND}{a}$, the same as for the revolving mirror. The latter method has, however, the same advantage over the former that the method of mirror and scale has over the direct reading of the needle of a galvanometer.

On the other hand, an important advantage for the method of the toothed wheel lies in the circumstance that the intensity of the return image is one-half of that which would appear if there were no toothed wheel, whereas with the revolving mirror this fraction is $\frac{n\beta}{rD}$ if the mirror has n facets), where β is the angular aperture of the concave mirror, and f is the focal length of the mirror, r is the distance from slit to revolving mirror, and D is the distance between stations.

In the actual experiments of Foucault, the greatest distance D was only 20 m (obtained by five reflections from concave mirrors), which, with a speed of five hundred turns per second, gives only 160″ for the angle 2α which is to be measured. The limit of accuracy of the method is about one second, so that under these conditions the results of Foucault's measurements can hardly be expected to be accurate to one part in one hundred and sixty. Foucault's result, 298,000, is in fact too small by this amount.[1]

In order to obtain a deflection 2α sufficiently large to measure with precision it is necessary to work with a much larger distance. The following plan renders this possible, and in a series of experiments (1878) the distance D was about 700 m and could have been made much greater.

The image-forming lens in the new arrangement is placed between the two mirrors, and (for maximum intensity of the return image) at a distance from the revolving mirror equal to the focal length of the lens. This necessitates a lens of long focus; for the radius of measurement r (from which α is determined by the relation $\delta = r \tan \alpha$, in which δ is the measured displacement of the image) is given by $r = \frac{f^2}{D}$, if f is the focal length of the lens; whence r is proportional to f^2. In the actual experiment,

[1] Apart from the mere matter of convenience in limiting the distance D to the insignificant 20 m (on account of the dimensions of the laboratory), it may be that this was in fact limited by the relative intensity of the return image as compared with that of the streak of light caused by the direct reflection from the revolving mirror, which in Foucault's experiments was doubtless superposed on the former. The intensity of the return image varies inversely as the cube of the distance, while that of the streak remains constant.

a non-achromatic lens of 25-m focus and 20-cm diameter was employed, and with this it was found that the intensity of the return light was quite sufficient even when the revolving mirror was far removed from the principal focus.

With so large a displacement, the inclined plane-parallel plate in the Foucault arrangement may be suppressed, the direct (real) image being observed. With 250 to 300 turns per second, a displacement of 100 to 150 mm was obtained which could be measured with an error of less than one ten-thousandth.

The measurement of D presents no serious difficulty. This was accomplished by means of a steel tape whose coefficient of stretch and of dilatation was carefully determined, and whose length under standard conditions was compared with a copy of the standard meter. The estimated probable error was of the order of 1:200,000.

The measurement of the speed of rotation presents some points of interest. The optical "beats" between the revolving mirror and an electrically maintained tuning fork were observed at the same time that the coincidence of the deflected image with the crosshairs of the eyepiece was maintained by hand regulation of an air blast which actuated the turbine attached to the revolving mirror. The number of vibrations of the fork *plus* the number of beats per second gives the number of revolutions per second in terms of the rate of the fork. This, however, cannot be relied upon except for a short interval, and it was compared before and after every measurement with a standard fork. This fork, whose temperature coefficient is well determined, is then compared, as follows, directly with a free pendulum.

For this purpose the pendulum is connected in series with a battery and the primary of an induction coil whose circuit is interrupted by means of a platinum knife edge attached to the pendulum passing through a globule of mercury. The secondary of the induction coil sends a flash through a vacuum tube, thus illuminating the edge of the fork and the crosshair of the observing microscope. If the fork makes an exact whole number (256) of vibrations during one swing of the pendulum, it appears at rest; but if there is a slight excess, the edge of the fork appears to execute a cycle of displacement at the rate of n per second. The rate of the fork is then $N \pm n$ per second of the free pendulum. This last is finally compared with a standard astronomical clock.[1] The order of accuracy is estimated as 1:200,000.

The final result of the mean of two such determinations of the velocity of light made under somewhat similar conditions but at a different time and locality is 299,895.

A determination of the velocity of light by a modification of the Foucault arrangement was completed by Newcomb in 1882. One of the essential improvements consisted in the use of a revolving steel prism with square section twice as long as wide. This permits the sending and receiving of the light on different parts of the mirror, thus eliminating the effect of direct reflection. It should also be mentioned that very accurate means were provided for measuring the deflection, and finally that the speed of the mirror was registered on a chronograph through a system of gears connected with the revolving mirror. Newcomb's result is 299,860.

[1] The average beat of such a clock may be extremely constant although the individual "seconds" vary considerably.

The original purpose of the Foucault arrangement was the testing of the question of the relative velocities of light in air and in water. For this purpose a tube filled with water and closed with plane-parallel glasses is interposed. There are then two return images of the source which would be superposed if the velocities were the same. By appropriately placed diaphragms these two images may be separated, and if there is any difference in velocities this is revealed by a relative displacement in the direction of rotation. This was found greater for the beam which had passed through the water column, and in which, therefore, the velocity must have been less. This result is in accordance with the undulatory theory and opposed to the corpuscular theory of light.

The experiments of Foucault do not appear to have shown more than qualitative results, and it should be of interest, not only to show that the velocity of light is less in water than in air, but that the ratio of the velocities is equal to the index of refraction of the liquid. Experiments were accordingly undertaken with water, the result obtained agreeing very nearly with the index of refraction. But on replacing the water by carbon disulphide, the ratio of velocities obtained was 1.75 instead of 1.64, the index of refraction. The difference is much too great to be attributed to errors of experiment.

Lord Rayleigh found the following explanation of the discrepancy. In the method of the toothed wheel the disturbances are propagated in the form of isolated groups of wave-trains. Rayleigh finds that the velocity of a group is not the same as that of the separate waves except in a medium without dispersion. The simplest form of group analytically considered is that produced by two simple harmonic wave-trains of slightly different frequencies and wave-lengths. Thus, let

$$y = \cos(nt - mx) + \cos(n_1 t - m_1 x),$$

in which $n = 2\pi/T$, and $m = 2\pi/\lambda$, T being the period and λ the wave-length. Let $n - n_1 = \partial n$, and $m - m_1 = \partial m$. Then

$$y = 2\cos\tfrac{1}{2}(\partial nt - \partial mx)\cos(nt - mx).$$

This represents a series of groups of waves such as illustrated in Figure 75.

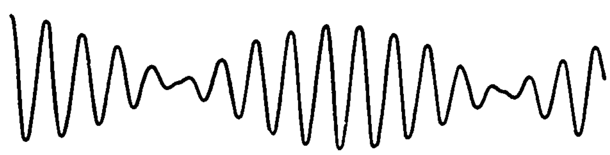


Fig. 75

The velocity of the waves is the ratio $V = n/m$, but the velocity of the group (e.g., the velocity of propagation of the maximum or the minimum) will be

$$V' = \partial n/\partial m,$$

or, since $n = mV$,

$$V' = \frac{\partial(mV)}{\partial m} = V + m\frac{\partial V}{\partial m} = V\left(1 + \frac{m\partial V}{V\partial m}\right),$$

or, since $m = 2\pi/\lambda$,

$$V' = V\left(1 - \frac{\lambda}{V}\frac{\partial V}{\partial \lambda}\right).$$

The demonstration is true, not only of this particular form of group, but (by the Fourier theorem) can be applied to a group of any form.

It is not quite so clear that this expression applies to the measurements made with the revolving mirror. Lord Rayleigh shows that in consequence of the Doppler effect there is a shortening of the waves at one edge of the beam of light reflected from the revolving mirror and a lengthening at the opposite edge, and since the velocity of propagation depends on the wave-length in a dispersive medium, there will be a rotation of the individual wave-fronts.

If ω is the angular velocity of the mirror, and ω_1 that of the dispersional rotation,

$$\omega_1=\frac{dV}{dy}=\frac{dV}{d\lambda}\frac{d\lambda}{dy},$$

where y is the distance from the axis of rotation. But

$$\frac{d\lambda}{dy}=2\omega\frac{\lambda}{V}\therefore\ \omega_1=2\omega\frac{\lambda}{V}\frac{dV}{d\lambda}.$$

The deflection actually observed is therefore

$$T(2\omega+\omega_1),$$

where T is the time required to travel distance $2D$; or

$$\frac{4D}{V}\omega\left(1+\frac{\lambda}{V}\frac{dV}{d\lambda}\right),$$

hence the velocity measured is

$$V''=V\div\left(1+\frac{\lambda}{V}\frac{dV}{d\lambda}\right),$$

or, to small quantities of the second order,

$$V''=V'=\text{group velocity}.$$ [1]

[1] J. W. Gibbs (*Nature*, 1886) shows that the measurement is in reality exactly that of groups and not merely an approximation.

The value of $\left(1+\frac{\lambda}{\mu}\frac{d\mu}{d\lambda}\right)$ for carbon disulphide for the mean wave-length of the visible spectrum is 0.93. Accordingly,

$$\frac{V_0}{V'}=\frac{V_0}{V}\frac{1}{0.93}=\frac{1.64}{0.93}=1.76\ ,$$

which agrees with the value found by experiment.

RECENT MEASUREMENTS OF THE VELOCITY OF LIGHT

In the expression for V, the velocity of light as determined by the revolving mirror, $V=4\pi ND/\alpha$, there are three quantities to be measured, namely, N, the speed of the mirror; D, the distance between stations; and α, the angular displacement of the mirror. As has already been mentioned, the values of N and D may be obtained to one part in one hundred thousand or less. But α cannot be measured to this order of accuracy. It has been pointed out by Newcomb[1] that this difficulty may be avoided by giving the revolving mirror a prismatic form and making the distance between the two stations so great that the return light is reflected at the same angle by the next following face of the prism

The following is an outline of a proposed attempt to realize such a project between Mount Wilson and Mount San Antonio near Pasadena, the distance being about 35 km. For this, given a speed of rotation of 1,060 turns per second, the angular displacement of the mirror during the double journey would be 90°; or, if the speed were half as great, an angle of 45° would suffice.[2] Accordingly,

[1] *Measures of the Velocity of Light.* Nautical Almanac Office, 1882.

[2] It may be noted that with eight surfaces the resulting intensity will be four times as great as with the revolving plane-parallel disk.

the revolving mirror may have the form of an octagon. It is, of course, very important that the angles should be equal, at least to the order of accuracy desired.

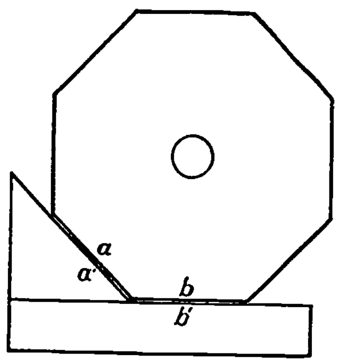


Fig. 76

This has already been attained as follows. The octagon, with faces polished and angles approximately correct, is applied to the test angle $a'b'$ made up of a 45° prism cemented to a true plane. The faces b_1b are made parallel by the interference fringes observed in monochromatic light. In general, the faces a_1a will not be parallel, and the angle between them is measured by the distance and inclination of the interference bands. The same process is repeated for each of the eight angles, and these are corrected by repolishing until the distance and inclination are the same for all, when the corresponding angles will also be equal. It has been found possible in this way to produce an octagon in which the average error was of the order of one-millionth, that is, about one-tenth to one-twentieth of a second.[1]

Another difficulty arises from the direct reflection and the scattered light from the revolving mirror. The former may be eliminated, as already mentioned, by slightly

[1] It may be noted that while a distortion may be expected when the mirror is in such rapid rotation, if the substance of the mirror (glass, in the present instance) is uniform, such distortion could only produce a very slight curvature and hence merely a minute change of focus.

inclining the revolving mirror, but to avoid the scattered light it is essential that the return ray be received on a different surface from the outgoing.

Again, in order to avoid the difficulty in maintaining the distant mirror perpendicular to the incident light, the return of the ray to the home station may be accom-

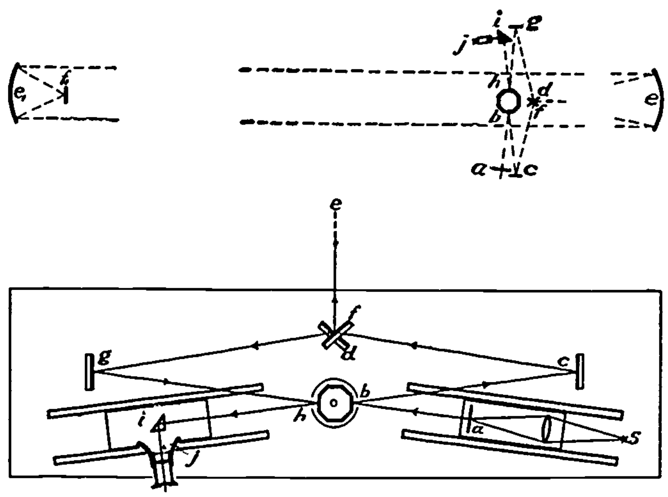


Fig. 77.—Light path $a, b, c, d, e, e_1, f_1, l_1, e, f, g, h, i, j$

plished exactly as in the Fizeau experiment, the only precaution required being the very accurate focusing of the beam on the small plane (better, concave) mirror at the focus of the distant collimator.

Finally, it is far less expensive to make both sending and receiving collimators silvered mirrors instead of lenses.

In Figure 77 is shown the arrangement of apparatus which fulfilled all these requirements.

Three determinations were undertaken between the home station at the Mount Wilson Observatory and Mount San Antonio 22 miles distant. The rate of the electric tuning fork was 132.25 vibrations per second, giving four stationary images of the revolving mirror when this was rotating at the rate of 529 turns per second. The fork was compared before and after every set of the observations with a free pendulum whose rate was found by comparison with an invar pendulum furnished and rated by the Coast and Geodetic Survey.

The result of eight measurements in 1924 gave

$$V_a = 299{,}735 \, .$$

Another series of observations with a direct comparison of the same electric fork with the Coast and Geodetic Survey pendulum[1] was completed in the summer of 1925 with a resulting value

$$V_a = 299{,}690 \, .$$

A third series of measurements was made in which the electric fork was replaced by a free fork making 528 vibrations per second maintained by an "audion circuit," thus insuring a much more nearly constant rate. The result of this measurement gave

$$V_a = 299{,}704 \, .$$

Giving these determinations the weights 1, 2, and 4, respectively, the result for the velocity in air is

$$V_a = 299{,}704 \, .$$

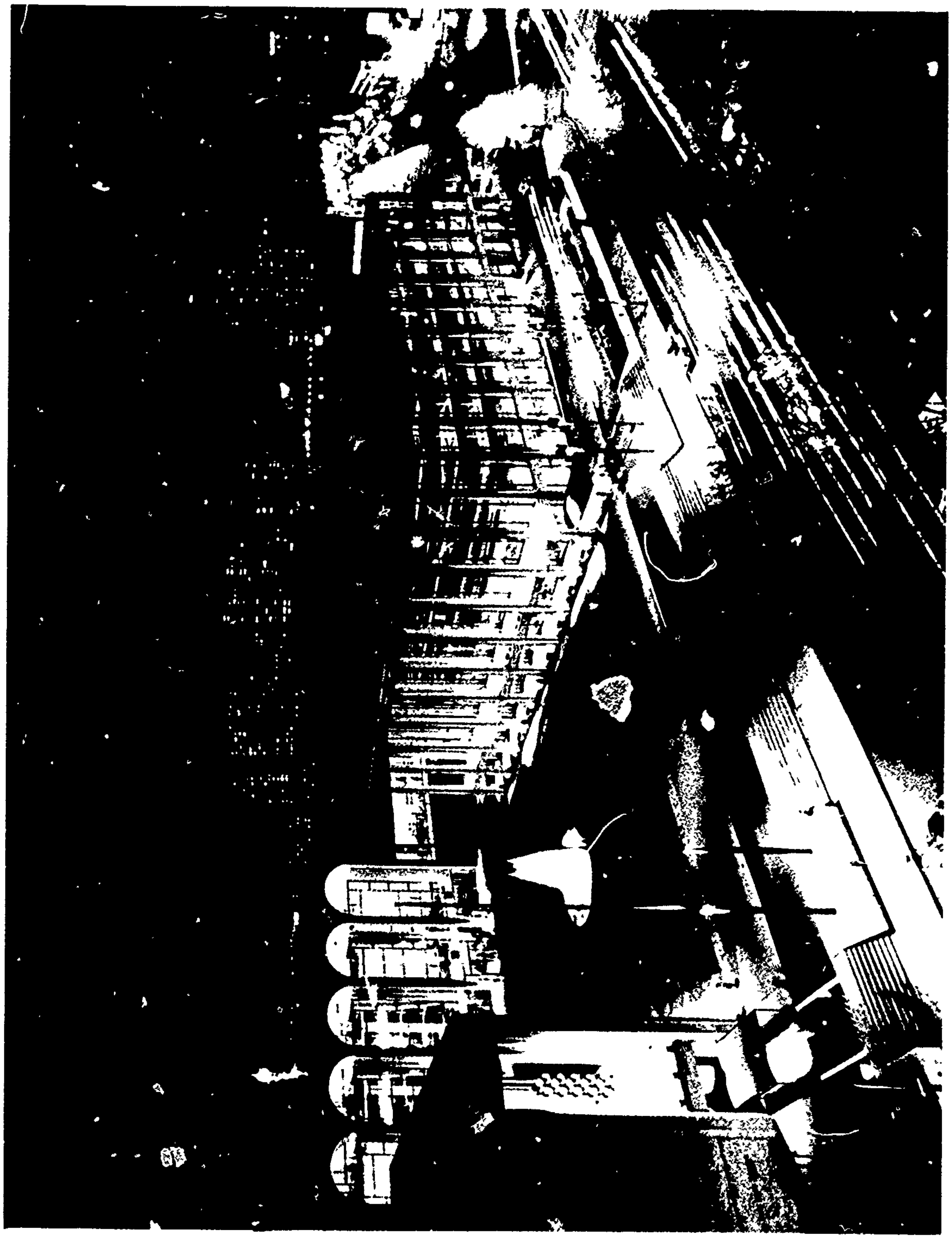
[1] This comparison was made by allowing the light from a very narrow slit to fall on a mirror attached to the pendulum. An image of the slit was formed by means of a good achromatic lens, in the plane of one edge of the fork, where it was observed by an ordinary eyepiece.

Table VIII shows the more reliable results of measurements of V with distance between stations, method used, and the weight assigned to each.

TABLE VIII

Author	D	Method	Wt.	V
Cornu	23 km	Toothed wheel	1	299,990
Perrotin.	12	Toothed wheel	1	299,900
M_1 and M_2.	0.6	Rev. mirror	1	299,880
Newcomb*. .	6.5	Rev. mirror	3	299,810
M_3.	35	Rev. mirror	5	299,800

* Newcomb's value omitting all discordant observations was 298,860.



Lincoln Center for the Performing Arts, New York City.

5 Popular Applications of Polarized Light

William A. Shurcliff and Stanley S. Ballard

A chapter from the book *Polarized Light* published in 1964.

If there is a logical order in which the various applications of polarizers and polarized light should be considered, the authors have never discovered it. The policy adopted here is to consider the most popular and "humanistic" applications first, and the more scientific and esoteric applications last.

POLARIZATION AND THE HUMAN EYE

The most humanistic fact about polarization of light is that it can be detected directly by the naked eye. Nearly anyone, if told carefully what to look for, can succeed in this. Sometimes he can even determine the form and azimuth of polarization.

What the observer actually "sees" is a certain faint pattern known as Haidinger's brush and illustrated in Fig. 10·1. The brush is so faint and ill-defined that it will escape notice unless the field of view is highly uniform: a clear blue sky makes an ideal background, and a brightly illuminated sheet of white paper is nearly as good. The best procedure for a beginner is to hold a linear polarizer in front of his eye, stare fixedly through it toward a clear blue sky, and, after five or ten seconds, suddenly turn the polarizer through 90°. Immediately the brush is seen. It fades away in two or three seconds, but reappears if the polarizer is again turned through 90°. The brush itself is sym-

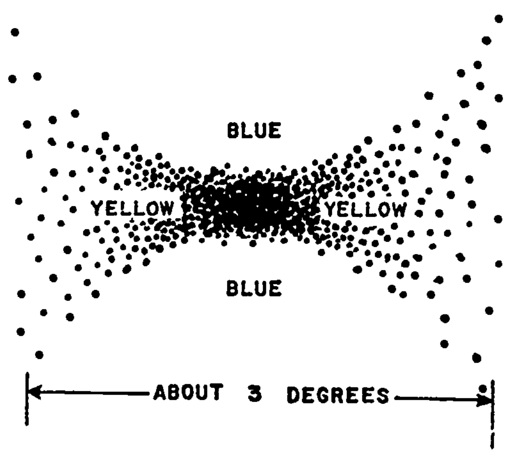


FIG. 10-1 Approximate appearance of Haidinger's brush when the vibration direction of the beam is vertical.

metric, double-ended, and yellow in color; it is small, subtending an angle of only about 2° or 3°. The adjacent areas appear blue, perhaps merely by contrast. The long axis of the brush is approximately perpendicular to the direction of electric vibration in the linearly polarized beam, i.e., perpendicular to the transmission axis of the polarizer used.

Circular polarization, too, can be detected directly by eye, and even the handedness can be determined. When an observer facing a clear blue sky places a right circular polarizer in front of his eye, he sees the yellow brush and finds that its long axis has an upward-to-the-right, downward-to-the-left direction, i.e., an azimuth of about +45°. This is true, of course, irrespective of the orientation of the polarizer, since a circle has no top or bottom. If he employs a *left* circular polarizer, he finds the brush to have a −45° orientation. In each case the pattern fades away rapidly, but can be restored to full vigor by switching to a polarizer of opposite handedness. Instead of using a circular polarizer the observer can use a single linear polarizer in series with a 90° retarder, the latter being held nearer to the eye. Turning the retarder through 90° reverses the handedness of the circular polarization.

Some people see the brush easily; others have difficulty. A few see the brush when looking innocently at the partially polarized blue sky, i.e., without using any polarizer at all, and even without meaning to see the brush. Some people see the brush more distinctly by linearly polarized light than by circularly polarized light, and for others the reverse is true. An observer may find the brush to have a slightly different orientation depending on which eye is used.

The spectral energy distribution of the light is important. If the light is rich in short-wavelength (blue) radiation, the brush is very noticeable, but if the short-wavelength radiation is eliminated by means of a yellow filter, the brush fails to appear. Use of a blue filter tends to accentuate the brush.

Although the phenomenon was discovered in 1844, by the Austrian mineralogist Haidinger, the cause is not yet fully understood. Presumably the thousands of tiny blue-light-absorbing bodies in the central (foveal) portion of the retina are dichroic and are oriented in a radial pattern, for example, a pattern such that the absorption axis of each body lies approximately along a radius from the center of the fovea. Incident linearly polarized light will then be absorbed more strongly in some parts of the pattern than in other parts and consequently some parts will fatigue more than others. When the vibration direction of the light is suddenly changed, the varying degrees of fatigue are revealed as a subjective radial pattern. Presumably no such dichroism or orientation pattern applies to longer wavelength (yellow and red) light; consequently a yellow sensation dominates in those regions where fatigue-to-blue has occurred.

The fact that circular polarization, also, may be detected perhaps implies that some transparent portion of the eye is weakly birefringent and acts like a retarder, converting circularly polarized light to linearly or elliptically polarized light. The direction of the major axis of the ellipse depends only on the direction of the fast axis of the retarding layer and hence remains fixed—unless the observer tips his head.

Perhaps physicists will some day write matrices to describe the retarding layers and dichroic areas of the eye. Poets were the first to see magic fire and jewels in the human eye; physicists will be the first to see matrices!

Bees, too, can detect the vibration direction of linearly polarized light. The experiments of the biologist K. von Frisch during World War II showed that bees "navigate" back and forth between hive and source of honey by using the sun as a guide. More interesting, when the sun is obscured by a large area of clouds the bees can still navigate successfully if they can see a bit of blue sky: they can detect the azimuth of linear polarization of the blue light and navigate with respect to it. One way of demonstrating the bee's ability to detect the azimuth of polarization is to place the bee in a large box the top of which consists of a huge sheet of linear polarizer, such as H-sheet. Each time the experimenter turns the polarizer to a different azimuth, the bee changes his direction of attempted travel correspondingly.

Certain other animals also can detect the polarization of skylight and navigate by it. This includes ants, beetles, and the fruit fly *Drosophila*. Probably many other examples will be discovered.

POLARIZATION OF SKY LIGHT

Blue-sky light traveling in a direction roughly at right angles to the sun's rays is partially polarized. When an observer holds a linear polarizer in front of his eye and gazes in a direction perpendicular to the direction of the sun, he finds that rotating the polarizer slowly causes the sky to change from bright to dark successively. The degree of polarization of sky light may reach 70 or 80 percent when the air is clear and dust-free, the sun is moderately low in the sky, and the observation direction is near the zenith.

The polarization is a result of the scattering of the sun's rays by the molecules in the air. Rayleigh's well-known inverse-fourth-power law relating scattering intensity to wavelength accounts for the blue color of the scattered light, and the asymmetry associated with the 90° viewing angle accounts for the polarization, as explained in Chapter 5. Some multiple scattering occurs, and this reduces the degree of polarization somewhat; when the observer ascends to a higher altitude, the amount of air involved is reduced, multiple scattering is reduced, and the degree of polarization is increased. A further increase results when a yellow or red filter is used to block the short-wavelength component of the light and transmit the long-wavelength component —the latter component is less subject to multiple scattering. (The situation is very different for infrared radiation of wavelength exceeding 2 microns: much of this radiation is produced by emission from the air itself, rather than by scattering, and this exhibits little or no polarization.)

Some persons are capable of detecting the polarization of sky light directly by eye, by virtue of the Haidinger brush phenomenon discussed in a preceding section; a few individuals find the brush noticeable enough to be a nuisance. Ordinarily, of course, it escapes notice and plays little part in the affairs of man. Its practical use by bees, ants, etc., has been indicated, and the importance to photographers is discussed in a later section.

POLARIZATION OF LIGHT UNDER WATER

A surprising fact about the polarization found in light present beneath the surface of the ocean (or of a pond) is that the predominant direction of electric vibration is horizontal. The opposite might be expected, since most of the light that enters the water enters *obliquely* from above, and the most strongly reflected component of obliquely incident light is the horizontally vibrating component. But oceanographers and biologists, working at depths of 5 to 30 feet in waters off Bermuda and in the Mediterranean Sea, have found the main cause of submarine polarization to be the scattering of the light by microscopic particles suspended in the water. Sunlight and sky light enter the water from above, and the average direction of illumination is roughly vertical; consequently the polarization form of the scattered light that travels horizontally toward an underwater observer is partially polarized with the electric vibration direction horizontal. The situation is much the same as that discussed in Chapter 5, except that the incident light has a more steeply downward direction and the asymmetric scattering is by microscopic particles instead of molecules.

Typically, the degree of polarization is 5 to 30 percent, an amount found to be important to a variety of underwater life. The water flea *Daphnia* tends to swim in a direction perpendicular to the electric vibration direction, for reasons not yet known. When tests are conducted in a tank filled with water that is free of suspended particles, so that the submarine illumination is practically unpolarized, *Daphnia* ceases to favor any one direction. But if suspended matter is added, thus restoring the polarization, *Daphnia* resumes the custom of traveling perpendicular to the vibration direction.

The arthropod *Limulus* (horse shoe crab) easily detects the polarization of the underwater light and is presumed to navigate with respect to the electric vibration direction. The same is true of the crustacean *Mysidium gracile* and various other forms of marine life. Most tend to swim perpendicularly to the vibration direction; some swim parallel to it; a few swim at different relative orientations depending on the time of day. For all of these animals, polarization is a compass that works even under water!

POLARIZING SUNGLASSES

The lenses of ordinary sunglasses employ absorbing materials that are isotropic, and accordingly the incident light is attenuated by a fixed factor irrespective of polarization form. This is unfortunate. The fact is that "glare" consists predominantly of light having a horizontal vibration direction. Why? For these reasons:

(a) The main source of light (sun and sky) is overhead, and consequently the main flux of light is downward.

(b) The surfaces that are most strongly illuminated by the downward flux are horizontal surfaces.

(c) Such surfaces are usually viewed obliquely, since a person seldom looks straight down.

(d) Most outdoor objects are of dielectric material.

(e) Light reflected obliquely from a horizontal dielectric surface is partially linearly polarized . ith the dominant vibration direction *horizontal,* as explained in Chapter 4.

Polarizing sunglasses take full advantage of this fact. The lenses are made of dichroic material (H-sheet, usually) oriented with the transmission axis vertical, as indicated in Fig. 10-2a, so

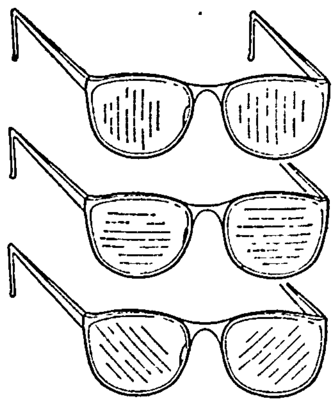


FIG. 10-2 Three types of polarizing spectacles. In (a) the transmission axis is vertical, for eliminating glare reflected from horizontal surfaces. In (b) the axis is horizontal, for eliminating reflections from vertical windows of trains, store-fronts (show-windows), etc. In (c) the axis directions are 45° and −45°, a standard arrangement used in viewing polarization-coded stereoscopic pictures.

that almost all of the horizontal vibrations are absorbed. The component having vertical vibration direction is transmitted. Usually some isotropic absorber is included in the lenses to absorb ultraviolet light strongly and blue and red light to a moderate extent; the sunglasses then have a greenish hue which has nothing to do with the polarization.

Motorists and vacationists find that polarizing sunglasses are helpful not only in reducing the brightness of the field of view as a whole, but also in enhancing the beauty of the scene. Because specularly reflected light is absorbed preferentially, roads, trees, grassy fields, etc., appear softer and more deeply colored through polarizers. Specularly reflected light tends to veil nature's inherent beauty; polarizing sunglasses remove the veil.

Fishermen and boatsmen enjoy another benefit from wearing polarizing sunglasses. They want to be able to see fish, rocks, etc., beneath the surface of the water, yet the light from such objects is dim and is usually lost in the "noise" of the sky light reflected obliquely from the surface. Since the reflected light is highly polarized with horizontal vibration direction, the polarizing sunglasses absorb this component strongly, and the visibility of the underwater objects is greatly increased. The increase is greatest when the viewing direction corresponds to the polarizing angle, which, for water, is about 53° from the normal. When the viewing direction is along the normal, i.e., straight down, there is no increase at all.

There is one interesting situation in which polarizing sunglasses produce little increase in visibility of underwater objects even when the angle of viewing is the polarizing angle. This situation occurs when the sky is clear and blue, the sun is low in the sky, and the pertinent portion of the sky is at 90° from the direction of the sun. Under these circumstances the light striking the water is already linearly polarized at such an azimuth that almost none of it is reflected. There is no task left for the sunglasses to perform—there is no reflected glare to suppress. The underwater objects are seen with great clarity. Persons unfamiliar with the polarization of sky light and with the dependence of oblique reflection on polarization form are likely to ascribe the remarkable clarity to "especially clear water" rather than to absence of reflection.

CAMERA FILTERS

Photographers often wish to enhance the contrast between blue sky and white clouds. Thirty years ago they did this by employing a yellow filter, which absorbed most of the blue light from the clear sky but transmitted most of the white light from the clouds. Using ordinary black-and-white film, they obtained excellent contrast by this method. Today, photographers are using color film increasingly, and the use of yellow filters is no longer permissible since it would eliminate all blue colors from the finished photograph.

The only known solution is to exploit the difference in polarization between blue sky and white clouds. Light from most portions of the blue sky is partially linearly polarized, as explained in a preceding section, and light from clouds is unpolarized. Therefore a neutral-color, linear polarizer mounted at the optimum azimuth in front of the lens will absorb a large fraction (e.g., 80 percent) of the sky light while transmitting a large fraction (nearly half) of the light from the clouds; thus the contrast is increased by a factor of two or three. The factor is less if the air is hazy, and more if the air is extremely clear (as in Arizona) and if the camera is aimed about 90° from the direction of the sun.

The usual way of choosing the azimuth of the polarizer is crude, but perhaps adequate. The photographer holds the polarizer in front of his eye, finds by trial and error which azimuth maximizes the contrast of the clouds in question, and then attempts to mount the polarizer on the camera without changing the azimuth of the polarizer. One type of polarizing filter for cameras is equipped with a small "satellite" polarizer mounted at the end of a short arm and aligned permanently with the main polarizer. The photographer installs the main polarizer in front of the lens, looks through the small polarizer and turns the arm to whatever azimuth maximizes the contrast. Both polarizers then have this optimum orientation. The satisfactoriness of the azimuth can be checked visually at any time. Instead of using these empirical methods, a scientifically minded photographer can proceed by dead reckoning, i.e., by following this well-known rule: Mount the polarizer so that its transmission axis lies in the plane determined by camera, sun, and object photographed. (So oriented, the polarizer performs a valuable additional service: it eliminates most of the specularly reflected light from trees, roads, etc., and enhances the softness and depth of color of the scene.)

When a photographer standing on a sidewalk tries to photograph objects situated behind a store window, the reflection of the street scene from the window may threaten to spoil the photograph. An excellent solution is to place the camera off to one side so that the window is seen obliquely at about the polarizing angle, and mount a linear polarizer in front of the lens; the polarizer is turned so that its transmission axis is horizontal, and the polarized light reflected from the window is absorbed. The authors have a friend who has applied this same principle to a pair of special spectacles he wears while touring the country by railroad. The lenses consist of polarizers oriented with the transmission axis horizontal, as indicated in Fig. 10-2b; thus when he gazes out of the train window in oblique forward direction, the reflected images of passengers and newspapers are wiped out, and the scenery appears in its pristine glory.

USE OF CIRCULAR POLARIZERS IN ELIMINATING PERPENDICULARLY REFLECTED LIGHT

Eliminating perpendicularly reflected lights is a different problem from that of eliminating obliquely reflected light. The process of oblique reflection at Brewster's angle causes the reflected beam to be linearly polarized, and accordingly a linear polarizer can eliminate the reflected beam entirely. But the process of *normal* reflection, i.e., with incident and reflected beams *perpendicular* to the smooth glossy surface in question, produces no polarization at all. How, then, can the specularly reflected light be eliminated while light originating behind the surface is transmitted freely?

The question is an important one to radar operators scanning the cathode-ray-oscilloscope screens on which dim greenish spots representing airborne objects appear. The screen proper is situated in a large evacuated tube, and the greenish light emerges through a curved glass window at the front end of the tube. (Sometimes the window is flat; sometimes a safety plate of glass or plastic is mounted close in front of it.) Often the operator has difficulty in seeing the greenish spots, not only because they are faint, but also because they may be masked by various extraneous images reflected by the front surface of the window, e.g., reflections of room lights and of people, clothing, papers, etc., situated near the operator. Extinguishing the room lights would eliminate these reflections, but would make it impossible for the operator to read instructions or make notes. What he needs is some kind of filter that will transmit the light originating behind the window and absorb the light reflected approximately perpendicularly from it.

This need is filled by the circular polarizer. Such a device, if mounted close in front of the window, will transmit nearly half of the light that originates behind the window, yet will eliminate about 99 percent of the room light that is reflected perpendicularly from it. The circular polarizer acts on the room light *twice:* it circularly polarizes room light that is approaching the window, then absorbs the reflected component. The logic behind this requires explanation. Two key facts must be kept in mind:

(1) A beam that is reflected perpendicularly and specularly by a smooth glossy surface has the same degree of polarization as the incident beam, since the reflection process does not introduce randomness of any kind.

(2) The reflection process reverses the handedness of polarization, because handedness is defined with respect to the propagation direction and the reflection process reverses the propagation direction.

If the polarizer is of right-circular type, as in the arrangement shown in Fig. 10-3, room light that passes through and ap-

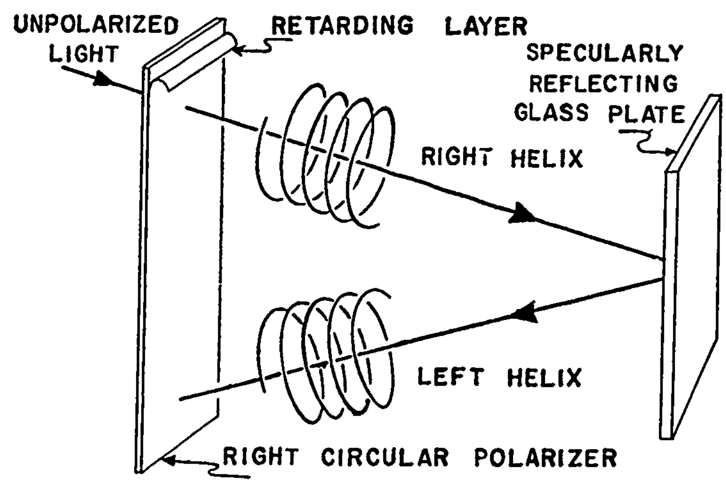


FIG. 10-3 Use of a circular polarizer in absorbing light reflected by a surface approximately perpendicular to the incident beam. Note that the reflection process reverses the handedness of circular polarization.

proaches the window is right-circularly polarized; the reflected light is *left*-circularly polarized and hence is *totally absorbed by the polarizer.* In effect, the circular polarizer "codes" the light, the window reverses the coding, and the polarizer then annihilates the reverse-coded beam. If both faces of the window are ideally flat and smooth, if the light is incident exactly along the normal, and if the polarizer is truly of circular type, the tion of the work.

It was also found that a trial with a much larger revolving mirror gave better definition, more light, and steadier speed of rotation; so that it seems probable that results of much greater accuracy may be obtained in a future investigation.

FINAL MEASUREMENTS

Observations with the same layout were resumed in the summer of 1926, but with an assortment of revolving mirrors.

The first of these was the same small octagonal glass mirror used in the preceding work. The result obtained this year was $V = 299{,}813$. Giving this a weight 2 and the result of preceding work weight 1 gives 299,799 for the weighted mean.

The other mirrors were a steel octagon, a glass 12-sider, a steel 12-sider, and a glass 16-sider.

The final results are summarized in Table VII.

TABLE VII

Turns per Second	Mirror	Number of Observations	Vel of Light *in Vacuo*
528	Glass oct.	576	299,797
528....	Steel oct.	195	299,795
352	Glass 12	270	299,796
352......	Steel 12	218	299,796
264.......	Glass 16	504	299,796

Weighted mean, 299,796 ± 1

reflected light is totally absorbed. Usually the conditions are less ideal: the rear surface of the window usually serves as support for the luminescent screen and has a matte appearance; the window is usually curved and much of the troublesome room light incident on the window makes an angle of 10° or 20° or more with the normal; and the polarizer, although circular with respect to some wavelengths, is elliptical with respect to others. Nevertheless, the improvement provided by the polarizer is large, and the amount of faint detail that the operator can see on the screen is greatly increased.

One precaution must be mentioned: reflections from the polarizer itself must be avoided. This is usually accomplished by tilting the polarizer forward so that the only reflected images the observer sees are images of a dark-colored floor or other dark objects.

Television sets, also, have been equipped successfully with circular polarizers. If the set is used in a brightly lit room, or is used outdoors, the circular polarizer performs a valuable service in trapping the specularly reflected glare and thus increasing the picture-vs-glare ratio by a factor of the order of 10.

VARIABLE-DENSITY FILTER

A pair of linear polarizers arranged in series is an almost ideal device for controlling the transmitted intensity of light. Rotating one polarizer through an angle θ with respect to the other causes used as windows of railroad cars and ocean liners. A person sitting near such a window turns a small knob to rotate one polarizer with respect to the other and thus reduce the intensity of the transmitted light to any extent desired.

One of the authors has experimented with a variable-density filter employing *three* linear polarizers in series, in order that a transmittance range of 10^8 to 1 could be achieved. The device worked well and, as expected, obeyed a cosine-fourth, rather than a cosine-square law.

THREE-DIMENSIONAL PHOTOGRAPHY AND THE USE OF POLARIZERS FOR CODING

Millions of polarizers found their way into the motion picture theaters of North America in 1952 and 1953 when stereoscopic (three-dimensional, or 3-D) movies achieved brief prominence. Each spectator wore a pair of polarizing spectacles called viewers, and polarizers were mounted in front of the projectors.

A photographer who enjoys looking at 3-D still pictures in his living room needs no polarizers. Usually he employs a small viewing box containing a light source and two lenses, one for each eye; a black partition, or septum, divides the box into right and left halves. The picture, consisting of two small transparencies mounted about two inches apart in a side-by-side arrangement on a cardboard frame, is inserted in the box so that the right-eye transparency lines up with the right lens and the left-eye transparency lines up with the left lens. (The two transparencies are, of course, slightly different because they were taken by cameras situated about two or three inches apart; the spacing used approximates the spacing of the two eyes.) The side-by-side arrangement of the two transparencies and the presence of the septum insure that the observer's right eye sees only the right transparency and the left eye sees only the left transparency. No cross-communication, or "cross-talk," can occur. Consequently the observer enjoys an impressively realistic stereoscopic illusion.

When 3-D motion-picture films are projected in a theater, many complications arise. Separate projectors must be provided for the right-eye and left-eye movie films, and the two projectors must be synchronized within about 0.01 second. Since there is just one large screen and this is to be viewed by hundreds of spectators, there can be no septum. Indeed, no practical geometrical method of preventing cross-talk is known.

Before the advent of mass-produced polarizers in the 1930's, an *analglyph* system of preventing cross-talk was invented. It applied wavelength coding to the two projected beams. The right-eye picture was projected through a long-wavelength (red) filter, and the left-eye picture was projected through a shorter-wavelength (green) filter. The spectator's viewers contained right and left lenses of red and green plastic, respectively, and accordingly each lens transmitted light from the appropriate projector and absorbed light from the other. Thus each eye received just the light intended for it. The system succeeded as a short-term novelty: stereoscopic illusions were created. But the system had two major defects: chromatic "retinal rivalry" between the two eyes, and incompatibility with the showing of colored motion pictures.

In the 1930's the problem was solved with éclat by a polarization-coding system, demonstrated with great impact at the New York World's Fair of 1939 and improved in later years. As indicated in Fig. 10-4, a linear polarizer oriented with its transmission axis at $-45°$ is placed in front of the projector used for the right-eye pictures, and a polarizer at $+45°$ is placed in front of the projector used for the left-eye pictures. Thus the two beams striking the movie screen are orthogonally coded. The lenses of the spectator's viewers consist of correspondingly oriented linear polarizers, and so each eye receives only light that originates in the appropriate projector. Superb stereoscopic illusions result. Since the polarizers perform well at all wavelengths in the visual range, color movies can be presented as easily and faithfully as can black-and-white movies.

The polarizers placed in front of the projectors consist, ordinarily, of K-sheet; as explained in Chapter 3, K-sheet is highly resistant to heat, and any polarizing filter placed close in front of a powerful projector is bound to heat up considerably since it necessarily absorbs about half the light. The lenses of the 3-D viewers are usually of HN-38 sheet; it has high major transmittance k_1 and small minor transmittance k_2, and it is inex-

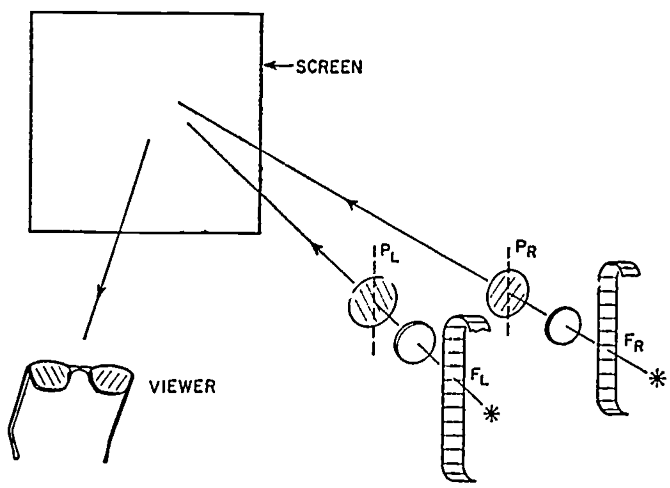


FIG. 10-4 Arrangement for projecting polarization-coded stereoscopic motion-picture films by means of two side-by-side projectors. Films F_R and F_L containing the "right-eye pictures" and "left-eye pictures" are mounted in the right and left projectors, which are equipped with linear polarizers P_R and P_L oriented at $-45°$ and $+45°$ respectively. The viewer contains correspondingly oriented polarizers, and accordingly each eye sees only the images intended for it.

pensive. The viewers are cheap enough (about 10¢ each) that they can be discarded after a single use.

The polarization-coding scheme has one limitation: if the spectator tilts his head to one side, the polarizers in his viewers no longer line up accurately with the respective polarizers on the projectors. Thus cross-talk occurs: the right eye sees faintly the image meant for the left eye, and vice versa: each eye sees a faint ghost image in addition to the main image. The spectator does not enjoy this. The difficulty could be avoided if the linear polarizers were replaced by high-quality, achromatic circular polarizers, but unfortunately no method is known for producing achromatic circular polarizers economically.

The effectiveness of any polarization-coding projection system is destroyed if the screen depolarizes the light appreciably. Screens that have a smooth aluminum coating usually conserve polarization to the extent of about 99 percent, but those having a matte white surface or a rough metallic coating produce much depolarization and hence much cross-talk between the two images. Many of the screens used in the innocent days of 1952 and 1953 were of the wrong type, and the resulting ghost images were a major annoyance. For that reason, and because of frequent lack of care in maintaining synchronism between the two projectors, movie-goers soon turned back to conventional 2-D pictures. Some nostalgia remains, however. Persons who were lucky enough to see a full-color, 3-D movie showing attractive actors filmed against a background of gorgeous scenery look forward to the time when well-made, well-presented 3-D movies, with their almost miraculous realism and intimacy, will animate the theaters once again.

THE VECTOGRAPH

The type of three-dimensional photography discussed in the preceding section is parallel-projected 3-D photography. The two motion-picture films are situated side-by-side, and two projectors are operated in parallel. During the late 1930's a radically new approach, called *vectography,* was developed by E. H. Land, J. Mahler, and others. In this system, the two films are arranged in series, bonded together. Because of the permanent series arrangement, many problems disappear. Only one projector is needed, and perfect synchronism is "guaranteed at the factory." Each pair of pictures (each vectograph) is projected as a single unit, in the same projector aperture and at the same time, and onto the same area of the same screen. If the film breaks, it can be spliced with no concern as to preservation of synchronism.

The method can succeed only if means are provided for preserving the identity of the two coincident projected beams. Again, polarization-coding is the answer. However, because the two images are bonded together in series, the coding must occur within the images themselves. In the system used by Land and Mahler each image consists of varying quantities of linearly dichroic molecules aligned in a common direction, and the directions employed in the two images are mutually at right angles. Dark areas in any one image contain a high concentration of dichroic molecules; light areas contain little or no dichroic material; but irrespective of concentration, the alignment direction is always the same. For the other image, the alignment direction is always orthogonal to the first. It is to be noted that the images contain no silver and no other isotropic absorber. Only aligned absorbers having high dichroic ratio are used.

A communications engineer would describe the vectograph by saying that it provides two distinct channels. Each is assigned to one image. Each is independent of the other. Since the vectograph images themselves perform the polarization coding, no polarizer is used in front of the projector; indeed, the interposition of such a filter would play havoc with the system. As before, the screen must preserve the polarization and the spectator's viewers must perform the appropriate decoding, or discriminating, act. Excellent stereoscopic effects are achieved. However, the production of vectograph film is a costly undertaking involving very specialized equipment, and constant attention is needed to maintain high enough dichroic ratio so that the channels are truly independent and ghost images are avoided.

Vectograph pictures of the "still" type are easier and cheaper to make than vectograph movies. Stereo pairs of aerial photographs of mountainous country, if presented in vectograph form, give a navigator (wearing an appropriate viewer) a very realistic impression of the terrain, and a map maker can prepare an accurate contour map from the vectograph with ease.

POLARIZING HEADLIGHTS

It is ironic that the main goal of Land and others in developing high-quality, large-area, low-cost polarizers has never been achieved. The polarizers are used with great success in dozens of applications, but not the application that was uppermost in the minds of the inventors.

Their goal was to eliminate glare from automobile headlights. In an era when dual-lane highways, circumferential bypasses, and other safety engineering advances were virtually unknown and the aim and focus of automobile headlights were highly erratic, the glare that confronted motorists at night was almost unbearable, and was an important cause of accidents. As early as 1920 several illumination engineers recognized that the glare could be eliminated by means of polarizers—if large-area polarizers could somehow be produced. If every headlight lens were covered by a linear polarizer oriented with the transmission axis horizontal and every windshield were covered with a linear polarizer oriented with its axis vertical, no direct light from the headlights of Car A could pass through the windshield of oncoming Car B. Drivers in both cars could see road-markings, pedestrians, and so forth, but neither would experience any glare from the other's headlights. Moreover, it would be permissible for each driver to use his *high* beam continuously, and accordingly his ability to see pedestrians, etc., would be greater than before, despite the fact that each polarizer would transmit only about half of the light incident on it.

It was soon recognized that the analyzing polarizer should not be made a permanent part of the windshield, but should be incorporated in a small visor situated just in front of the driver's eyes. During the day, when headlights were not in use, the visor could be swung out of the way. It was also recognized that care should be taken to make sure the headlight polarizers had sufficient light-leak, i.e., sufficiently large k_2 value, that the headlights of oncoming cars would not disappear entirely!

Land and his colleagues moved rapidly. They invented a whole series of polarizers, each superior to its predecessor. The first successful type, J-sheet, employed aligned, microscopic crystals of the dichroic mineral herapathite; the method of manufacture is described in Chapter 3. Then came H-sheet, which was better in nearly every respect and in addition was easier to make. Finally, K-sheet appeared; it had most of the superb qualities of the earlier materials and the added virtue of being unaffected by fairly high temperature, such as 215°F. To persons seeking polarizers for use in headlights, K-sheet appeared to be the pot of gold at the end of a polarized rainbow.

Concurrently, several better ways of orienting the polarizers were proposed. One attractive scheme was to orient the headlight polarizers and the visor polarizer at the identical azimuth, namely −45°, as indicated in Fig. 10-5. Then, even a polarization-conserving object in the path of the headlights would appear

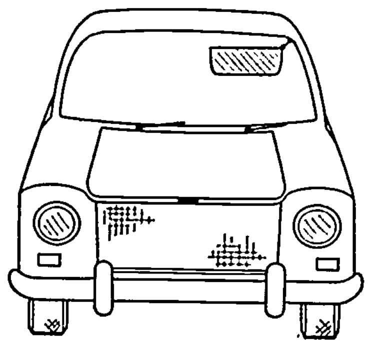


FIG. 10-5 Automobile equipped with headlight polarizers and a visor polarizer oriented at −45°. When two such cars approach one another, each driver is protected from the glare from the headlights of the other.

to the driver to be brightly illuminated. (This would not be the case if his visor polarizer were crossed with his headlight polarizers.) The −45° system disposed of the headlight glare problem adequately: if two cars A and B both equipped in this manner approached one another at night, each driver's visor would be crossed with the other car's headlight polarizers, and neither driver would experience any glare.

Using the Mueller calculus, Billings and Land compared a wide variety of polarizer orientation schemes, and found several to be particularly attractive. Perhaps the best system was one called "−55°, −35°." The transmission axes of the headlight polarizers and visor polarizer are at 55° and 35° from the vertical, respectively, an arrangement that minimizes complications stemming from the obliquity of the portion of the windshield situated just in front of the driver.

Despite the successes on all technical fronts, the project bogged down. To this day no one knows just why. Probably many little reasons were responsible. Among these were the following:

(1) The polarizers absorbed slightly more than half of the light incident on them, and accordingly the automobile manufacturers felt that they would have to increase the power of the lamps themselves and perhaps use larger generators and batteries also.

(2) Some windshields were moderately birefringent; therefore

they would act like retarders, alter the polarization form of the incident light, and allow some glare to leak through.

(3) Nearly every year the automobile manufacturers increased the backward tilt of the windshields; such tilt tends to alter the polarization form of light having an oblique vibration direction, and hence leads to glare-leak.

(4) Passengers, as well as drivers, would require visors, since passengers also dislike glare.

(5) Pedestrians might find that the glare was worse than ever, unless they too employed polarizing visors or spectacles.

(6) The system would succeed only if adopted by *all* car manufacturers, and therefore no one manufacturer would gain any promotional advantage from it.

(7) The first few drivers to put the system to use would get little benefit from it for at least a year or two, i.e., until millions of other cars were similarly equipped.

(8) It was difficult to decide when and how to force the owners of old cars to install the necessary polarizers on their cars.

(9) The patents on the only fully satisfactory polarizers were held by a single company.

(10) To introduce the system would require formal, coordinated action by all States.

(11) Improvements in headlight design and aiming, the increasing numbers of dual-lane highways, and the brighter street lamps used in cities and suburbs led some people to believe that the need for a polarization-type of glare control was no longer acute.

However, persons who have actually experienced the polarization method of glare removal are convinced that the drawbacks are trivial compared to the benefits.

Perhaps some day the system will be tried out on a pilot scale in a small, isolated community, where all the cars could be equipped with polarizers in a few weeks. Perhaps an island of moderate size would make a good test ground. If the system is found to be highly successful there, it will presumably spread throughout every country that teems with automobiles.

Action at a Distance

James Clerk Maxwell

A scientific paper published in 1873.

I HAVE no new discovery to bring before you this evening. I must ask you to go over very old ground, and to turn your attention to a question which has been raised again and again ever since men began to think.

The question is that of the transmission of force. We see that two bodies at a distance from each other exert a mutual influence on each other's motion. Does this mutual action depend on the existence of some third thing, some medium of communication, occupying the space between the bodies, or do the bodies act on each other immediately, without the intervention of anything else?

The mode in which Faraday was accustomed to look at phenomena of this kind differs from that adopted by many other modern inquirers, and my special aim will be to enable you to place yourselves at Faraday's point of view, and to point out the scientific value of that conception of *lines of force* which, in his hands, became the key to the science of electricity.

When we observe one body acting on another at a distance, before we assume that this action is direct and immediate, we generally inquire whether there is any material connection between the two bodies; and if we find strings, or rods, or mechanism of any kind, capable of accounting for the observed action between the bodies, we prefer to explain the action by means of these intermediate connections, rather than to admit the notion of direct action at a distance.

Thus, when we ring a bell by means of a wire, the successive parts of the wire are first tightened and then moved, till at last the bell is rung at a distance by a process in which all the intermediate particles of the wire have taken part one after the other. We may ring a bell at a distance in other ways, as by forcing air into a long tube, at the other end of which is a cylinder with a piston which is made to fly out and strike the bell. We may also use a wire; but instead of pulling it, we may connect it at one end with a voltaic battery, and at the other with an electro-magnet, and thus ring the bell by electricity.

Here are three different ways of ringing a bell. They all agree, however, in the circumstance that between the ringer and the bell there is an unbroken line of communication, and that at every point of this line some physical process goes on by which the action is transmitted from one end to the other. The process of transmission is not instantaneous, but gradual; so that there is an interval of time after the impulse has been given to one extremity of the line of communication, during which the impulse is on its way, but has not reached the other end.

It is clear, therefore, that in many cases the action between bodies at a distance may be accounted for by a series of actions between each successive pair of a series of bodies which occupy the intermediate space; and it is asked, by the advocates of mediate action, whether, in those cases in which we cannot perceive the intermediate agency, it is not more philosophical to admit the existence of a medium which we cannot at present perceive, than to assert that a body can act at a place where it is not.

To a person ignorant of the properties of air, the transmission of force by means of that invisible medium would appear as unaccountable as any other example of action at a distance, and yet in this case we can explain the whole process, and determine the rate at which the action is passed on from one portion to another of the medium.

Why then should we not admit that the familiar mode of communicating motion by pushing and pulling with our hands is the type and exemplification of all action between bodies, even in cases in which we can observe nothing between the bodies which appears to take part in the action?

Here for instance is a kind of attraction with which Professor Guthrie has made us familiar. A disk is set in vibration, and is then brought near a light suspended body, which immediately begins to move towards the disk, as if drawn towards it by an invisible cord. What is this cord? Sir W. Thomson has pointed out that in a moving fluid the pressure is least where the velocity is greatest. The velocity of the vibratory motion of the air is greatest nearest the disk. Hence the pressure of the air on the suspended body is less on the side nearest the disk than on the opposite side, the body yields to the greater pressure, and moves toward the disk.

The disk, therefore, does not act where it is not. It sets the air next it in motion by pushing it, this motion is communicated to more and more distant portions of the air in turn, and thus the pressures on opposite sides of the suspended body are rendered unequal, and it moves towards the disk in consequence of the excess of pressure. The force is therefore a force of the old school—a case of *vis a tergo*—a shove from behind.

The advocates of the doctrine of action at a distance, however, have not been put to silence by such arguments. What right, say they, have we to assert that a body cannot act where it is not? Do we not see an instance of action at a distance in the case of a magnet, which acts on another magnet not only at a distance, but with the most complete indifference to the nature of the matter which occupies the intervening space? If the action depends on something occupying the space between the two magnets, it cannot surely be a matter of indifference whether this space is filled with air or not, or whether wood, glass, or copper, be placed between the magnets.

Besides this, Newton's law of gravitation, which every astronomical observation only tends to establish more firmly, asserts not only that the heavenly bodies act on one another across immense intervals of space, but that two portions of matter, the one buried a thousand miles deep in the interior of the earth, and the other a hundred thousand miles deep in the body of the sun, act on one another with precisely the same force as if the strata beneath which each is buried had been non-existent. If any medium takes part in transmitting this action, it must surely make some difference whether the space between the bodies contains nothing but this medium, or whether it is occupied by the dense matter of the earth or of the sun.

But the advocates of direct action at a distance are not content with instances of this kind, in which the phenomena, even at first sight, appear to favour their doctrine. They push their operations into the enemy's camp, and maintain that even when the action is apparently the pressure of contiguous portions of matter, the contiguity is only apparent—that a space *always* intervenes between the bodies which act on each other. They assert, in short, that so far from action at a distance being impossible, it is the only kind of action which ever occurs, and that the favourite old *vis a tergo* of the schools has no existence in nature, and exists only in the imagination of schoolmen.

The best way to prove that when one body pushes another it does not touch it, is to measure the distance between them. Here are two glass lenses, one of which is pressed against the other by means of a weight. By means of the electric light we may obtain on the screen an image of the place where the one lens presses against the other. A series of coloured rings is formed on the screen. These rings were first observed and first explained by Newton. The particular colour of any ring depends on the distance between the surfaces of the pieces of glass. Newton formed a table of the colours corresponding to different distances, so that by comparing the colour of any ring with Newton's table, we may ascertain the distance between the surfaces at that ring. The colours are arranged in rings because the surfaces are spherical, and therefore the interval between the surfaces depends on the distance from the line joining the centres of the spheres. The central spot of the rings indicates the place where the lenses are nearest together, and each successive ring corresponds to an increase of about the 4000th part of a millimètre in the distance of the surfaces.

The lenses are now pressed together with a force equal to the weight of an ounce; but there is still a measurable interval between them, even at the place where they are nearest together. They are not in optical contact. To prove this, I apply a greater weight. A new colour appears at the central spot, and the diameters of all the rings increase. This shews that the surfaces are now nearer than at first, but they are not yet in optical contact, for if they were, the central spot would be black. I therefore increase the weights, so as to press the lenses into optical contact.

But what we call optical contact is not real contact. Optical contact indicates only that the distance between the surfaces is much less than a wave-length of light. To shew that the surfaces are not in real contact, I remove the weights. The rings contract, and several of them vanish at the centre. Now it is possible to bring two pieces of glass so close together, that they will not tend to separate at all, but adhere together so firmly, that when torn asunder the glass will break, not at the surface of contact, but at some other place. The glasses must then be many degrees nearer than when in mere optical contact.

Thus we have shewn that bodies begin to press against each other whilst still at a measurable distance, and that even when pressed together with great force they are not in absolute contact, but may be brought nearer still, and that by many degrees

Why, then, say the advocates of direct action, should we continue to maintain the doctrine, founded only on the rough experience of a pre-scientific age, that matter cannot act where it is not, instead of admitting that all the facts from which our ancestors concluded that contact is essential to action were in reality cases of action at a distance, the distance being too small to be measured by their imperfect means of observation?

If we are ever to discover the laws of nature, we must do so by obtaining the most accurate acquaintance with the facts of nature, and not by dressing up in philosophical language the loose opinions of men who had no knowledge of the facts which throw most light on these laws. And as for those who introduce ætherial, or other media, to account for these actions, without any direct evidence of the existence of such media, or any clear understanding of how the media do their work, and who fill all space three and four times over with æthers of different sorts, why the less these men talk about their philosophical scruples about admitting action at a distance the better.

If the progress of science were regulated by Newton's first law of motion, it would be easy to cultivate opinions in advance of the age. We should only have to compare the science of to-day with that of fifty years ago; and by producing, in the geometrical sense, the line of progress, we should obtain the science of fifty years hence.

The progress of science in Newton's time consisted in getting rid of the celestial machinery with which generations of astronomers had encumbered the heavens, and thus "sweeping cobwebs off the sky."

Though the planets had already got rid of their crystal spheres, they were still swimming in the vortices of Descartes. Magnets were surrounded by effluvia, and electrified bodies by atmospheres, the properties of which resembled in no respect those of ordinary effluvia and atmospheres.

When Newton demonstrated that the force which acts on each of the heavenly bodies depends on its relative position with respect to the other bodies, the new theory met with violent opposition from the advanced philosophers of the day, who described the doctrine of gravitation as a return to the exploded method of explaining everything by occult causes, attractive virtues, and the like.

Newton himself, with that wise moderation which is characteristic of all his speculations, answered that he made no pretence of explaining the mechanism by which the heavenly bodies act on each other. To determine the mode in which their mutual action depends on their relative position was a great step in science, and this step Newton asserted that he had made. To explain the process by which this action is effected was a quite distinct step, and this step Newton, in his *Principia*, does not attempt to make.

But so far was Newton from asserting that bodies really do act on one another at a distance, independently of anything between them, that in a letter to Bentley, which has been quoted by Faraday in this place, he says:—

"It is inconceivable that inanimate brute matter should, without the mediation of something else, which is not material, operate upon and aff ✝ other matter without mutual contact, as it must do if gravitation, in the sense of Epicurus, be essential and inherent in it......That gravity should be innate, inherent, and essential to matter, so that one body can act upon another at a distance, through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it."

Accordingly, we find in his *Optical Queries*, and in his letters to Boyle, that Newton had very early made the attempt to account for gravitation by means of the pressure of a medium, and that the reason he did not publish these investigations "proceeded from hence only, that he found he was not able, from experiment and observation, to give a satisfactory account of this medium, and the manner of its operation in producing the chief phenomena of nature*."

The doctrine of direct action at a distance cannot claim for its author the discoverer of universal gravitation. It was first asserted by Roger Cotes, in his preface to the *Principia*, which he edited during Newton's life. According to Cotes, it is by experience that we learn that all bodies gravitate. We do not learn in any other way that they are extended, movable, or solid. Gravitation, therefore, has as much right to be considered an essential property of matter as extension, mobility, or impenetrability.

And when the Newtonian philosophy gained ground in Europe, it was the opinion of Cotes rather than that of Newton that became most prevalent, till at last Boscovich propounded his theory, that matter is a congeries of mathematical points, each endowed with the power of attracting or repelling the others according to fixed laws. In his world, matter is unextended, and contact

* Maclaurin's *Account of Newton's Discoveries*.

is impossible. He did not forget, however, to endow his mathematical points with inertia. In this some of the modern representatives of his school have thought that he "had not quite got so far as the strict modern view of 'matter' as being but an expression for modes or manifestations of 'force'*."

But if we leave out of account for the present the development of the ideas of science, and confine our attention to the extension of its boundaries, we shall see that it was most essential that Newton's method should be extended to every branch of science to which it was applicable—that we should investigate the forces with which bodies act on each other in the first place, before attempting to explain *how* that force is transmitted. No men could be better fitted to apply themselves exclusively to the first part of the problem, than those who considered the second part quite unnecessary.

Accordingly Cavendish, Coulomb, and Poisson, the founders of the exact sciences of electricity and magnetism, paid no regard to those old notions of "magnetic effluvia" and "electric atmospheres," which had been put forth in the previous century, but turned their undivided attention to the determination of the law of force, according to which electrified and magnetized bodies attract or repel each other. In this way the true laws of these actions were discovered, and this was done by men who never doubted that the action took place at a distance, without the intervention of any medium, and who would have regarded the discovery of such a medium as complicating rather than as explaining the undoubted phenomena of attraction.

We have now arrived at the great discovery by Örsted of the connection between electricity and magnetism. Örsted found that an electric current acts on a magnetic pole, but that it neither attracts it nor repels it, but causes it to move round the current. He expressed this by saying that "the electric conflict acts in a revolving manner."

The most obvious deduction from this new fact was that the action of the current on the magnet is not a push-and-pull force, but a rotatory force, and accordingly many minds were set a-speculating on vortices and streams of æther whirling round the current.

But Ampère, by a combination of mathematical skill with experimental ingenuity, first proved that two electric currents act on one another, and then analysed this action into the resultant of a system of push-and-pull forces between the elementary parts of these currents.

* Review of Mrs Somerville, *Saturday Review*, Feb. 13, 1869.

The formula of Ampère, however, is of extreme complexity, as compared with Newton's law of gravitation, and many attempts have been made to resolve it into something of greater apparent simplicity.

I have no wish to lead you into a discussion of any of these attempts to improve a mathematical formula. Let us turn to the independent method of investigation employed by Faraday in those researches in electricity and magnetism which have made this Institution one of the most venerable shrines of science.

No man ever more conscientiously and systematically laboured to improve all his powers of mind than did Faraday from the very beginning of his scientific career. But whereas the general course of scientific method then consisted in the application of the ideas of mathematics and astronomy to each new investigation in turn, Faraday seems to have had no opportunity of acquiring a technical knowledge of mathematics, and his knowledge of astronomy was mainly derived from books.

Hence, though he had a profound respect for the great discovery of Newton, he regarded the attraction of gravitation as a sort of sacred mystery, which, as he was not an astronomer, he had no right to gainsay or to doubt, his duty being to believe it in the exact form in which it was delivered to him. Such a dead faith was not likely to lead him to explain new phenomena by means of direct attractions.

Besides this, the treatises of Poisson and Ampère are of so technical a form, that to derive any assistance from them the student must have been thoroughly trained in mathematics, and it is very doubtful if such a training can be begun with advantage in mature years.

Thus Faraday, with his penetrating intellect, his devotion to science, and his opportunities for experiments, was debarred from following the course of thought which had led to the achievements of the French philosophers, and was obliged to explain the phenomena to himself by means of a symbolism which he could understand, instead of adopting what had hitherto been the only tongue of the learned.

This new symbolism consisted of those lines of force extending themselves in every direction from electrified and magnetic bodies, which Faraday in his mind's eye saw as distinctly as the solid bodies from which they emanated.

The idea of lines of force and their exhibition by means of iron filings was nothing new. They had been observed repeatedly, and investigated mathematically as an interesting curiosity of science. But let us hear Faraday himself, as he introduces to his reader the method which in his hands became so powerful*.

"It would be a voluntary and unnecessary abandonment of most valuable aid if an experimentalist, who chooses to consider magnetic power as represented by lines of magnetic force, were to deny himself the use of iron filings. By their employment he may make many conditions of the power, even in complicated cases, visible to the eye at once, may trace the varying direction of the lines of force and determine the relative polarity, may observe in which direction the power is increasing or diminishing, and in complex systems may determine the neutral points, or places where there is neither polarity nor power, even when they occur in the midst of powerful magnets. By their use probable results may be seen at once, and many a valuable suggestion gained for future leading experiments."

Experiment on Lines of Force.

In this experiment each filing becomes a little magnet. The poles of opposite names belonging to different filings attract each other and stick together, and more filings attach themselves to the exposed poles, that is, to the ends of the row of filings. In this way the filings, instead of forming a confused system of dots over the paper, draw together, filing to filing, till long fibres of filings are formed, which indicate by their direction the lines of force in every part of the field.

The mathematicians saw in this experiment nothing but a method of exhibiting at one view the direction in different places of the resultant of two forces, one directed to each pole of the magnet; a somewhat complicated result of the simple law of force.

But Faraday, by a series of steps as remarkable for their geometrical definiteness as for their speculative ingenuity, imparted to his conception of these lines of force a clearness and precision far in advance of that with which the mathematicians could then invest their own formulæ.

In the first place, Faraday's lines of force are not to be considered merely as individuals, but as forming a system, drawn in space in a definite manner

* *Exp. Res.* 3284.

so that the number of the lines which pass through an area, say of one square inch, indicates the intensity of the force acting through the area. Thus the lines of force become definite in number. The strength of a magnetic pole is measured by the number of lines which proceed from it; the electro-tonic state of a circuit is measured by the number of lines which pass through it.

In the second place, each individual line has a continuous existence in space and time. When a piece of steel becomes a magnet, or when an electric current begins to flow, the lines of force do not start into existence each in its own place, but as the strength increases new lines are developed within the magnet or current, and gradually grow outwards, so that the whole system expands from within, like Newton's rings in our former experiment. Thus every line of force preserves its identity during the whole course of its existence, though its shape and size may be altered to any extent.

I have no time to describe the methods by which every question relating to the forces acting on magnets or on currents, or to the induction of currents in conducting circuits, may be solved by the consideration of Faraday's lines of force. In this place they can never be forgotten. By means of this new symbolism, Faraday defined with mathematical precision the whole theory of electro-magnetism, in language free from mathematical technicalities, and applicable to the most complicated as well as the simplest cases. But Faraday did not stop here. He went on from the conception of geometrical lines of force to that of physical lines of force. He observed that the motion which the magnetic or electric force tends to produce is invariably such as to shorten the lines of force and to allow them to spread out laterally from each other. He thus perceived in the medium a state of stress, consisting of a tension, like that of a rope, in the direction of the lines of force, combined with a pressure in all directions at right angles to them.

This is quite a new conception of action at a distance, reducing it to a phenomenon of the same kind as that action at a distance which is exerted by means of the tension of ropes and the pressure of rods. When the muscles of our bodies are excited by that stimulus which we are able in some unknown way to apply to them, the fibres tend to shorten themselves and at the same time to expand laterally. A state of stress is produced in the muscle, and the limb moves. This explanation of muscular action is by no means complete. It gives no account of the cause of the excitement of the state of stress, nor does it even investigate those forces of cohesion which enable the muscles to support this stress. Nevertheless, the simple fact, that it substitutes a kind of action which extends continuously along a material substance for one of which we know only a cause and an effect at a distance from each other, induces us to accept it as a real addition to our knowledge of animal mechanics.

For similar reasons we may regard Faraday's conception of a state of stress in the electro-magnetic field as a method of explaining action at a distance by means of the continuous transmission of force, even though we do not know how the state of stress is produced.

But one of Faraday's most pregnant discoveries, that of the magnetic rotation of polarised light, enables us to proceed a step farther. The phenomenon, when analysed into its simplest elements, may be described thus:— Of two circularly polarised rays of light, precisely similar in configuration, but rotating in opposite directions, that ray is propagated with the greater velocity which rotates in the same direction as the electricity of the magnetizing current.

It follows from this, as Sir W. Thomson has shewn by strict dynamical reasoning, that the medium when under the action of magnetic force must be in a state of rotation--that is to say, that small portions of the medium, which we may call molecular vortices, are rotating, each on its own axis, the direction of this axis being that of the magnetic force.

Here, then, we have an explanation of the tendency of the lines of magnetic force to spread out laterally and to shorten themselves. It arises from the centrifugal force of the molecular vortices.

The mode in which electromotive force acts in starting and stopping the vortices is more abstruse, though it is of course consistent with dynamical principles.

We have thus found that there are several different kinds of work to be done by the electro-magnetic medium if it exists. We have also seen that magnetism has an intimate relation to light, and we know that there is a theory of light which supposes it to consist of the vibrations of a medium. How is this luminiferous medium related to our electro-magnetic medium?

It fortunately happens that electro-magnetic measurements have been made from which we can calculate by dynamical principles the velocity of propagation of small magnetic disturbances in the supposed electro-magnetic medium.

This velocity is very great, from 288 to 314 millions of metres per second, according to different experiments. Now the velocity of light, according to

Foucault's experiments, is 298 millions of metres per second. In fact, the different determinations of either velocity differ from each other more than the estimated velocity of light does from the estimated velocity of propagation of small electro-magnetic disturbance. But if the luminiferous and the electro-magnetic media occupy the same place, and transmit disturbances with the same velocity, what reason have we to distinguish the one from the other? By considering them as the same, we avoid at least the reproach of filling space twice over with different kinds of æther.

Besides this, the only kind of electro-magnetic disturbances which can be propagated through a non-conducting medium is a disturbance transverse to the direction of propagation, agreeing in this respect with what we know of that disturbance which we call light. Hence, for all we know, light also may be an electro-magnetic disturbance in a non-conducting medium. If we admit this, the electro-magnetic theory of light will agree in every respect with the undulatory theory, and the work of Thomas Young and Fresnel will be established on a firmer basis than ever, when joined with that of Cavendish and Coulomb by the key-stone of the combined sciences of light and electricity—Faraday's great discovery of the electro-magnetic rotation of light.

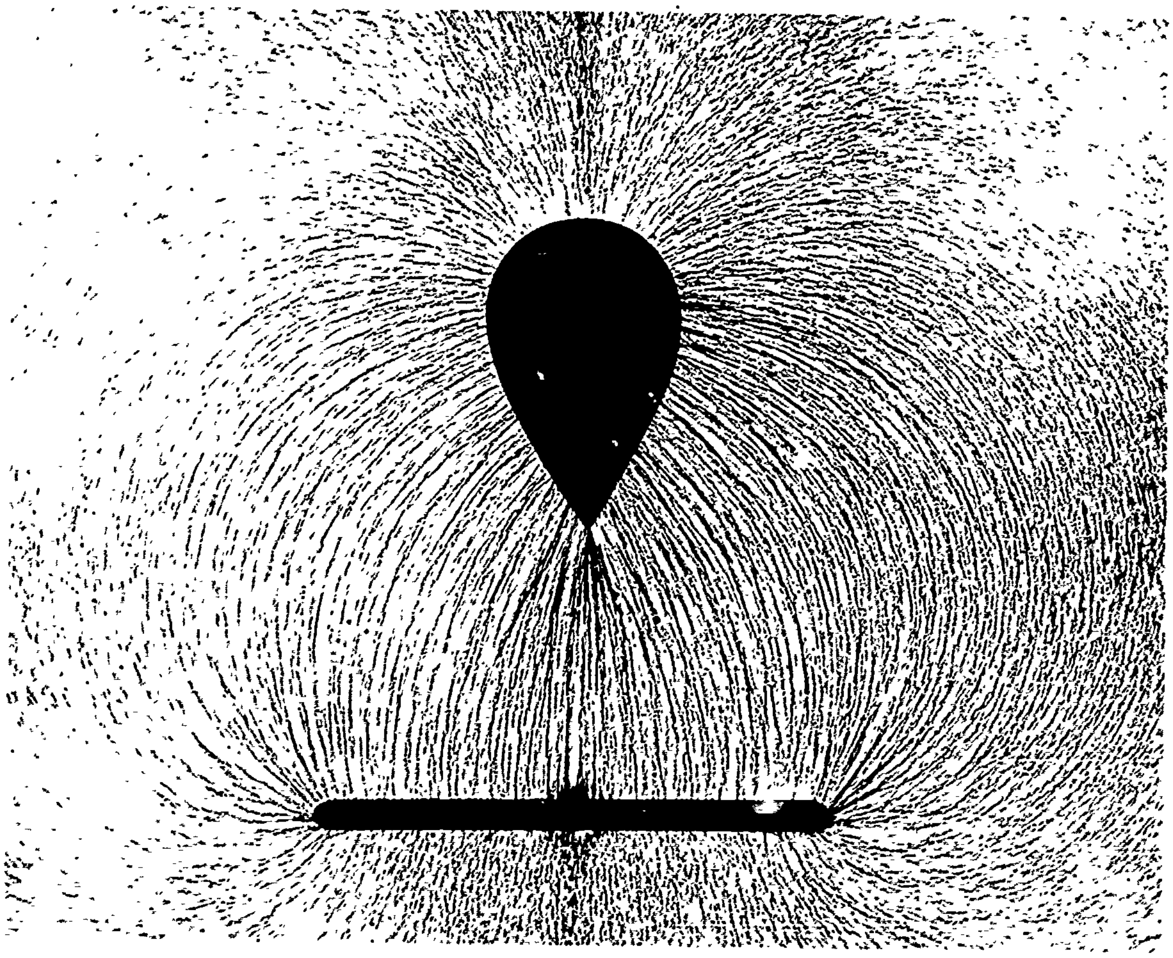
The vast interplanetary and interstellar regions will no longer be regarded as waste places in the universe, which the Creator has not seen fit to fill with the symbols of the manifold order of His kingdom. We shall find them to be already full of this wonderful medium; so full, that no human power can remove it from the smallest portion of space, or produce the slightest flaw in its infinite continuity. It extends unbroken from star to star; and when a molecule of hydrogen vibrates in the dog-star, the medium receives the impulses of these vibrations; and after carrying them in its immense bosom for three years, delivers them in due course, regular order, and full tale into the spectroscope of Mr Huggins, at Tulse Hill.

But the medium has other functions and operations besides bearing light from man to man, and from world to world, and giving evidence of the absolute unity of the metric system of the universe. Its minute parts may have rotatory as well as vibratory motions, and the axes of rotation form those lines of magnetic force which extend in unbroken continuity into regions which no eye has seen, and which, by their action on our magnets, are telling us in language not yet interpreted, what is going on in the hidden underworld from minute to minute and from century to century.

And these lines must not be regarded as mere mathematical abstractions. They are the directions in which the medium is exerting a tension like that of a rope, or rather, like that of our own muscles. The tension of the medium in the direction of the earth's magnetic force is in this country one grain weight on eight square feet. In some of Dr Joule's experiments, the medium has exerted a tension of 200 lbs. weight per square inch.

But the medium, in virtue of the very same elasticity by which it is able to transmit the undulations of light, is also able to act as a spring. When properly wound up, it exerts a tension, different from the magnetic tension, by which it draws oppositely electrified bodies together, produces effects through the length of telegraph wires, and when of sufficient intensity, leads to the rupture and explosion called lightning.

These are some of the already discovered properties of that which has often been called vacuum, or nothing at all. They enable us to resolve several kinds of action at a distance into actions between contiguous parts of a continuous substance. Whether this resolution is of the nature of explication or complication, I must leave to the metaphysicians.



Electric field between oppositely charged conductors.

A brief, informal review of the electronic age, past and present.

7 The Electronic Revolution

Arthur C. Clarke

1962

The electron is the smallest thing in the universe; it would take thirty thousand million, million, million, million of them to make a single ounce. Yet this utterly invisible, all but weightless object has given us powers over nature of which our ancestors never dreamed. The electron is our most ubiquitous slave; without its aid, our civilization would collapse in a moment, and humanity would revert to scattered bands of starving, isolated savages.

We started to use the electron fifty years before we discovered it. The first practical application of electricity (which is nothing more than the ordered movement of electrons) began with the introduction of the telegraph in the 1840's. With really astonishing speed, a copper cobweb of wires and cables spread across the face of the world, and the abolition of distance had begun. For over a century we have taken the instantaneous transfer of news completely for granted; it is very hard to believe that when Lincoln was born, communications were little faster than in the days of Julius Caesar.

Although the beginning of "electronics" is usually dated around the 1920's, this represents a myopic view of technology. With the hindsight of historical perspective, we can now see that the telegraph and the telephone are the first two landmarks of the electronic age. After Alexander Graham Bell had sent his voice from one room to another in 1876, society could never be the same again. For the telephone was the first electronic device to enter the home and to affect directly the lives of ordinary men and women, giving them the almost godlike power of projecting their personalities and thoughts from point to point with the speed of lightning.

Until the closing years of the nineteenth century, men used and handled electricity without knowing what it was, but in the 1890's they began to investigate its fundamental nature, by observing what happened when an electric current was passed through gases at very low pressures. One of the first, and most dramatic, results of this work was the invention of the X-ray tube, which may be regarded as the ancestor of all the millions of vacuum tubes which followed it. A cynic might also argue that it is the only electronic device wholly beneficial to mankind—though when it was invented many terrified spinsters, misunderstanding its powers, denounced poor Röntgen as a violator of privacy.

There is an important lesson to be learned from the X-ray tube. If a scientist of the late Victorian era had been asked "In what way could money best be spent to further the progress of medicine?" he would never by any stretch of the imagination have replied: "By encouraging research on the conduction of electricity through rarefied gases." Yet that is what would have been the right answer, for until the discovery of X rays doctors and surgeons were like blind men, groping in the dark. One can never predict the outcome of fundamental scientific research, or guess what remote and unexpected fields of knowledge it will illuminate.

X rays were discovered in 1895—the electron itself just one year later. It was then realized that an electric current consists of myriads of these submicroscopic particles, each carrying a minute negative charge. When a current flows through a solid conductor such as a piece of copper wire, we may imagine the electrons creeping like grains of sand through the interstices between the (relatively) boulder-sized copper atoms. Any individual electron does not move very far, or very fast, but it jostles its neighbor and so the impulse travels down the line at speeds of thousands of miles a second. Thus when we switch on a light, or send a Morse dash across a transatlantic cable, the response at the other end is virtually instantaneous.

But electrons can also travel *without* wires to guide them, when they shoot across the empty space of a vacuum tube like a hail of machine-gun bullets. Under these conditions, no longer entangled in solid matter, they are very sensitive to the pull and tug of electric fields, and as a result can be used to amplify faint signals. You demonstrate the principle involved every time you hold a hose-pipe in your hand; the slightest movement of your wrist produces a much greater effect at the far end of the jet. Something rather similar happens to the beam of electrons crossing the space in a vacuum tube; they can thus multiply a millionfold the feeble impulses picked up by a radio antenna, or paint a fluorescent picture on the end of a television screen.

Until 1948, electronics was almost synonymous with the vacuum tube. The entire development of radio, talkies, radar, television, long-distance telephony, up to that date depended upon little glass bottles containing intricate structures of wire and mica. By the late 1940's the vacuum tube had shrunk from an object as large as (and sometimes almost as luminous as) an electric light bulb, to a cylinder not much bigger than a man's thumb. Then three scientists at the Bell Telephone Laboratories invented the transistor and we moved from the Paleoelectronic to the Neoelectronic Age.

Though the transistor is so small—its heart is a piece of crystal about the size of a rice grain—it does everything that a radio tube can do. However, it requires only a fraction of the power and space, and is potentially much more reliable. Indeed, it is hard to see how a properly designed transistor can ever wear out; think of little Vanguard I, still beeping away up there in space, and liable to continue indefinitely until some exasperated astronaut scoops it up with a butterfly net.

The transistor is of such overwhelming importance because it (and its still smaller successors) makes practical hundreds of electronic devices which were previously too bulky, too expensive or too unreliable for everyday use. The pocket radio is a notorious example; whether we like it or not, it points the way inevitably to a day when person-to-person communication is universal. Then everyone in the world will have his individual telephone number, perhaps given to him at birth and serving all the other needs of an increasingly complex society (driving license, social security, credit card, permit to have additional children, etc.). You may not know where on Earth your friend Joe Smith may be at any particular moment; but you will be able to dial him instantly—if only you can remember whether his number is 8296765043 or 8296756043.

Obviously, there are both advantages and disadvantages in such a "personalized" communication system; the solitude which we all need at some time in our lives will join the vanished silences of the pre-jet age. Against this, there is no other way in which a really well-informed *and* fast-reacting democratic society can be achieved on the original Greek plan—with direct participation of every citizen in the affairs of the state. The organization of such a society, with feedback in both directions from the humblest citizen to the President of the World, is a fascinating exercise in political planning. As usual, it is an exercise that will not be completed by the time we need the answers.

A really efficient and universal communications system, giving high-quality reception on all bands between all points on the Earth, can be achieved only with the aid of satellites. As they come into general use, providing enormous information-handling capacity on a global basis, today's patterns of business, education, entertainment, international affairs will change out of all recognition. Men will be able to meet face to face (individually, or in groups) without ever leaving their homes, by means of closed circuit television. As a result of this, the enormous amount of commuting and traveling that now takes place from home to office, from ministry to United

Nations, from university to conference hall will steadily decrease. There are administrators, scientists and businessmen today who spend about a third of their working lives either traveling or preparing to travel. Much of this is stimulating, but most of it is unnecessary and exhausting.

The improvement of communications will also render obsolete the city's historic role as a meeting place for minds and a center of social intercourse. This is just as well anyway, since within another generation most of our cities will be strangled to death by their own traffic.

But though electronics will ultimately separate men from their jobs, so that (thanks to remote manipulation devices) not even a brain surgeon need be within five thousand miles of his patient, it must also be recognized that few of today's jobs will survive long into the electronic age. It is now a cliché that we are entering the Second Industrial Revolution, which involves the mechanization not of energy, but of thought. Like all clichés this is so true that we seldom stop to analyze what it means.

It means nothing less than this: There are no routine, noncreative activities of the human mind which cannot be carried out by suitably designed machines. The development of computers to supervise industrial processes, commercial transactions and even military operations has demonstrated this beyond doubt. Yet today's computers are morons compared to those that they themselves are now helping to design.

I would not care to predict how many of today's professions will survive a hundred years from now. What happened to the buggywhip makers, the crossing sweepers, the scriveners, the stonebreakers of yesteryear? (I mention the last because I can just remember them, hammering away at piles of rock in the country lanes of my childhood.) Most of our present occupations will follow these into oblivion, as the transistor inherits the earth.

For as computers become smaller, cheaper and more reliable they will move into every field of human activity. Today they are in the office; tomorrow they will be in the home. Indeed, some very simple-minded computers already do our household chores; the device that programs a washing machine to perform a certain sequence of operations is a specialized mechanical brain. Less specialized ones would be able to carry out almost all the routine operations in a suitably designed house.

Because we have so many more pressing problems on our hands, only the science-fiction writers—those trail-blazers of the future—have given much thought to the social life of the later electronic age. How will our descendants be educated for leisure, when the working week is only a few hours? We have already seen, on a worldwide scale, the cancerous growths resulting from idleness and lack of usable skills. At every street corner in a great city you will find lounging groups of leather-jacketed, general-purpose bioelectric computers of a performance it will take us centuries and trillions of dollars to match. What is their future—and ours?

More than half a century ago H. G. Wells described, in *The Time Machine*, a world of decadent pleasure lovers, bereft of goals and ambitions, sustained by subterranean machines. He set his fantasy eight hundred thousand years in the future, but we may reach a similar state of affairs within a dozen generations. No one who contemplates the rising curve of technology from the Pilgrim fathers to the Apollo Project dare deny that this is not merely possible, but probable.

For most of history, men have been producers; in a very few centuries, they will have to switch to the role of consumers, devoting their energies 100 per cent to absorbing the astronomical output of the automated mines, farms and factories.

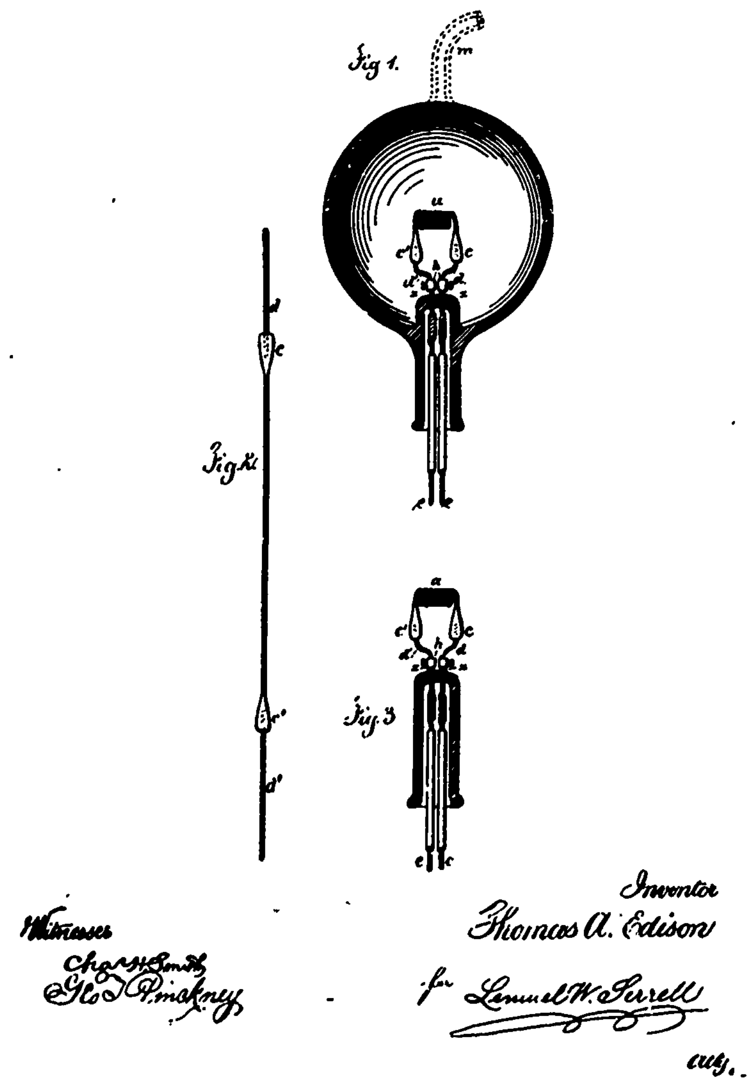
Does this *really* matter, since only a tiny fraction of the human race has ever contributed to artistic creation, scientific discovery or philosophical thought, which in the long run are the only significant activities of mankind? Archimedes and Aristotle, one cannot help thinking, would still have left their marks on history even if they had lived in a society based on robots instead of human slaves. In any culture, they would be consumers of goods, but producers of thought.

We should not take too much comfort from this. The electronic computers of today are like the subhuman primates of ten million years ago, who could have given any visiting Martians only the faintest hints of their potentialities, which included the above mentioned Archimedes and Aristotle. Evolution is swifter now; electronic intelligence is only decades, not millions of years, ahead.

And *that*—not transistor radios, automatic homes, global TV—is the ultimate goal of the Electronic Revolution. Whether we like it or not, we are on a road where there is no turning back; and waiting at its end are our successors.

T. A. EDISON.
Electric-Lamp.

No. 223,898. Patented Jan. 27, 1880.



EDISON'S PATENT on the incandescent lamp was accompanied by this drawing. The labeled parts are the carbon filament (*a*), thickened ends of filament (*c*), platinum wires (*d*), clamps (*h*), leading wires (*x*), copper wires (*e*), tube to vacuum pump (*m*).

The Invention of the Electric Light

Matthew Josephson

1959

"I can hire mathematicians, but mathematicians can't hire me!" By such declarations in the time of his success and world-wide fame Thomas Alva Edison helped to paint his own portrait as an authentic American folk hero: the unlettered tinkerer and trial-and-error inventor who achieved his results by persistence and a native knack for things. He is said, for example, to have tried more than 1,600 kinds of material ("paper and cloth, thread, fishline, fiber, celluloid, boxwood, coconut-shells, spruce, hickory, hay, maple shavings, rosewood, punk, cork, flax, bamboo and the hair out of a red-headed Scotchman's beard") until he hit upon the loop of carbonized cotton thread that glowed in a vacuum for more than half a day on October 21, 1879 Today, in a world that relies for its artificial illumination largely on his incandescent lamp, this invention is not regarded as an especially profound contribution to technology. It rates rather as a lucky contrivance of Edison's cut-and-try methods—of a piece with his stock ticker, mimeograph machine, phonograph and alkaline storage-battery—in the esteem of a public that has come to appreciate the enormous practical significance of higher mathematics and abstruse physical theory.

If Edison's contribution to the light of the world consisted solely in the selection of a filament, this estimate of his person and achievements might be allowed to stand. But the history that is so obscured by legend tells quite another story. Edison's electric light was not merely a lamp but a system of electric lighting. His invention was an idea rather than a thing. It involved not only technology but also sociology and economics. Edison was indisputably the first to recognize that electric lighting would require that electricity be generated and distributed at high voltage in order to subdivide it among a great many high-resistance "burners," each converting current at low amperage (that is, in small volume) with great efficiency into light.

In the 15 months between the time he conceived his invention and the date on which he demonstrated it to the public, Edison and his associates designed and built a new type of electric generator, successfully adapted the then much-scorned parallel or "multiple-arc" circuit that would permit individual lights to be turned on or off separately and, last of all, fashioned a lamp to meet the specifications of his system. The laboratory notebooks of those months of frantic labor show the Wizard of Menlo Park endowed with all the prodigious capacities attributed to him by contemporary legend. They show in addition that this self-taught technologist was possessed of a profound grasp of the nature of electricity and an intuitive command of its logic and power.

It was on September 8, 1878, that Edison was inspired to devote his talents full time to the challenge of electric lighting. On that day he went to Ansonia, Conn., to visit the brass-manufacturing plant of William Wallace, co-inventor with Moses G. Farmer of the first practical electric dynamo in the U. S. Wallace showed Edison eight brilliant carbon-arc lights of 500 candle-power each, powered by a dynamo of eight horsepower. It was with such a system that Wallace and Farmer, as well as Charles Brush of Cleveland, were then beginning to introduce the electric light on a commercial scale, for street-lighting and for illuminating factories and shops. Farmer had made the first demonstration of arc-lighting in this country two years earlier, at the Centennial Exposition in Philadelphia, and John Wanamaker's store in that city was already illuminated with arc lights.

Carbon arcs are still employed in searchlights and in theater floodlights and projectors to produce light of high intensity The current crossing a small gap between the electrodes creates an arc. Ionization and oxidation of the carbon in the heat of the arc generate a brilliant blue-white light.

In the 1870's Europe was a decade ahead of the U. S. in the technology of arc-lighting. Stores, railway stations, streets and lighthouses in Britain and France were equipped with arc lights. Shedding an almost blinding glare, they burned in open globes that emitted noxious gases, and they could be employed only high overhead on streets or in public buildings. Since they consumed large amounts of current, they had to be wired in series, that is, connected one to another in a single continuous circuit so that all had to be turned on or off together. The multiple-arc circuit, with the lights connected as in the rungs of a ladder between the main leads of the circuit, was not adapted to such systems and was considered prohibitive in cost.

Edison himself had experimented with arc lights, using carbon strips as burners. He had also investigated the

EDITOR'S NOTE

The author has based this article on material in his biography *Edison*, just published by the McGraw-Hill Book Company. Copyright © 1959 by Matthew Josephson.

incandescent light, as had many inventors before him. But the slender rod or pencil of carbon or metal would always burn up, sooner rather than later, upon being heated to incandescence by the current. It would do so though substantially all of the air had been pumped out of the glass envelope in which it was contained. Edison had abandoned the effort to devote himself to a more promising invention, the phonograph.

Now at Wallace's establishment, confronted with the achievements of others in the field, he regained his earlier enthusiasm. As an eyewitness recalled, "Edison was enraptured. ... He fairly gloated. He ran from the instruments [the dynamos] to the lights, and then again from the lights back to the electric instruments. He sprawled over a table and made all sorts of calculations. He calculated the power of the instruments and the lights, the probable loss of power in transmission, the amount of coal the instrument would use in a day, a week, a month, a year."

To William Wallace he said challengingly, "I believe I can beat you making the electric light. I do not think you are working in the right direction." They shook hands in friendly fashion, and with a diamond-pointed stylus Edison signed his name and the date on a goblet provided by his host at dinner.

From Edison's own complete and explicit notebooks and from the buoyant interviews that he gave to the press at this time we know what made him feel in such fine fettle as he left Wallace's plant. "I saw for the first time everything in practical operation," he said. "I saw the thing had not gone so far but that I had a chance. The intense light had not been subdivided so that it could be brought into private houses. In all electric lights theretofore obtained the light was very great, and the quantity [of lights] very low. I came home and made experiments two nights in succession. I discovered the necessary secret, so simple that a bootblack might understand it. . . . The subdivision of light is all right."

The Subdivision of Light

At this time there flashed into Edison's mind the image of the urban gas-lighting system, with its central gashouse and gas mains running to smaller branch pipes and leading into many dwelling places at last to gas jets that could be turned on or off at will. During the past half-century gas-lighting had reached the stature of a major industry in the U. S. It was restricted, of course, to the cities; three fourths of the U. S. population still lived in rural areas by the dim glow of kerosene lamps or candles. Ruminating in solitude, Edison sought to give a clear statement to his objective. In his notebook, under the title "Electricity versus Gas as a General Illuminant," he wrote:

"Object. E . . . to effect exact imitation of all done by gas, to replace lighting by gas by lighting by electricity. To improve the illumination to such an extent as to meet all requirements of natural, artificial and commercial conditions. . . . Edison's great effort—not to make a large light or a blinding light, but a small light having the mildness of gas."

To a reporter for one of the leading New York dailies who had shadowed him to Ansonia, Edison described a vision of a central station for electric lighting that he would create for all of New York City. A network of electric wire would deliver current for a myriad of small household lights, unlike the dazzling arc lights made by Farmer and Brush. In some way electric current would be metered and sold. Edison said he hoped to have his electric-light invention ready in six weeks! At Menlo Park, N.J., where his already famous workshop was located, he would wire all the residences for light and hold a "grand exhibition."

Thus from the beginning Edison riveted his attention not so much upon the search for an improved type of incandescent filament as upon the analysis of the social and economic conditions for which his invention was intended. As he turned with immense energy to expanding the facilities at Menlo Park and securing the essential financing, he continued his studies of the gas-lighting industry. In parallel he projected the economics of the electric-lighting system he envisioned.

Gas had its inconvenience and dangers. "So unpleasant . . . that in the new Madison Square theater every gas jet is ventilated by small tubes to carry away the products of combustion." But whatever is to replace gas must have "a general system of distribution—the only possible means of economical illumination." Gathering all the back files of the gas industry's journals and scores of volumes bearing on gas illumination, he studied the operations and habits of the industry, its seasonal curves and the layout of its distribution systems. In his mind he mapped out a network of electric-light lines for an entire city, making the shrewd judgment: "Poorest district for light, best for power—thus evening up whole city." He meant that in slum districts there would be higher demand for small industrial motors. Against tables for the cost of converting coal to gas he calculated the cost of converting coal and steam into electric energy. An expert gas engineer, whose services Edison engaged at this time, observed that few men knew more about the world's gas business than did Edison.

Edison had a *homo oeconomicus* within him, a well-developed social and commercial sense, though he was careless of money and was not an accountant of the type exemplified by his contemporary John D. Rockefeller. Before the experimental work on his invention was under way, he had formed a clear notion, stated in economic terms, of what its object must be. This concept guided his search and determined the pattern of his technical decisions, so that the result would be no scientific toy but a product useful to people everywhere. By his initial calculation of the capital investment in machinery and copper for a whole system of light distribution he was led to define the kind of light he sought and the kind of generating and distributing system he needed.

Backers of the Electric Light

In the crucial matter of financing his inventive work Edison had the generous and imaginative aid of Grosvenor Lowrey, a patent and corporation lawyer well established in the financial community of Wall Street. Lowrey had fallen completely under Edison's spell and regarded him much as a collector of paintings regards a great artist whose works he believes are destined for immortality. Using his extensive connections and the favorable press-notices that he encouraged Edison to secure during late September and early October, 1878, Lowrey assembled a sponsoring syndicate of some of the most important financiers of the time. The underwriters of the Edison Electric Light Company, which was incorporated in mid-October, included William H. Vanderbilt and J. P. Morgan's partner Egisto Fabbri. This was an unprecedented development in U. S. business. Inventors had been backed in the development of inventions already achieved; Edison's financiers were backing him in research that was to lead to a hoped-for invention. In many respects the venture marks the beginning in this country of close relations between finance and technology.

"Their money," Edison said, "was in-

EDISON AND HIS PHONOGRAPH were photographed in 1878 by Mathew Brady. He had worked with electric lights but had turned to the more promising phonograph. In the year that this photograph was made, however, he resumed his work on lighting.

MENLO PARK was depicted in *Frank Leslie's Illustrated Newspaper* for January 10, 1880. The barnlike "tabernacle" of Edison's laboratory is visible at the far right. In its windows passengers on the nearby railroad could see his experimental lights burning.

vested in confidence of my ability to bring it back again." The 31-year-old Edison was by now a well-known figure in Wall Street. His quadruplex telegraph system, by which four separate messages could be transmitted over a single wire, had furnished the pivotal issue in the vast economic war waged between Western Union and the rival telegraph empire of the robber baron Jay Gould. Edison's carbon microphone had transformed the telephone from an instrument of limited usefulness to an efficient system of long-range communication that was now radiating across the country. The shares of gas-lighting enterprises had tumbled on the New York and London exchanges upon Edison's announcement, in the press campaign instigated by Lowrey, that he was now about to displace gas with electricity in the lighting of homes and factories.

The alliance between Edison and his sponsors was nonetheless an uneasy one. The first rift appeared before the end of October, when the rival inventor William Sawyer and his partner Albon Man announced that they had "beaten" Edison and applied for a patent on a carbon-pencil light in a nitrogen-filled glass tube. There was a flutter of panic in the directorate of the Edison Electric Light Company. The suggestion was made that Edison should join forces with Sawyer and Man. Lowrey passed the suggestion on to S. L. Griffin, a former junior executive at Western Union whom Lowrey had hired to help Edison with his business affairs.

Griffin sent back a hasty "confidential" reply: "I spoke to Mr. Edison regarding the Sawyer-Man electric light. . . . I was astonished at the manner in which Mr. Edison received the information. He was visibly agitated and said it was the old story, that is, lack of confidence. . . . No combination, no consolidation for him. I do not feel at liberty to repeat all he said, but I do feel impelled to suggest respectfully that as little be said to him as possible with regard to the matter."

In view of Edison's talent for candid and salty language Griffin's reticence is understandable. After that there was no further talk of consolidation with Sawyer or any other inventor.

The Menlo Park Laboratory

In his belief that he would "get ahead of the other fellows" Edison was sustained by his unbounded confidence in his laboratory, its superior equipment and its staff. The Menlo Park laboratory was still the only full-time industrial research organization in the country, in itself perhaps Edison's most important invention. During this period the physical plant was greatly expanded; a separate office and library, a house for two 80-horsepower steam engines, and a glass blower's shed were added to the original barnlike "tabernacle." Even more important, Edison had collected a nucleus of talented engineers and skilled craftsmen, who were of inestimable help to him in working out his ideas.

The self-taught Edison thought primarily in concrete, visual terms. When he was at work on the quadruplex telegraph, he had even built a model made up of pipes and valves corresponding to the wires and relays of his system, and with running water replacing the electric current, so that he could actually see how it worked. But now he would have to depend far more on theory and mathematics.

One of the happiest effects of Grosvenor Lowrey's personal influence was the hiring of Francis R. Upton, a young electrical engineer who had worked for a year in the Berlin laboratory of the great physicist Hermann von Helmholtz. Edison jocularly nicknamed Upton "Culture," and, according to an oft-told story, put the "green" mathematician in his place with one of his scientific practical jokes. He brought out a pear-shaped glass lamp-bulb and gave it to Upton, asking him to calculate its content in cubic centimeters. Upton drew the shape of the bulb exactly on paper, and derived from this an equation for the bulb's volume. He was about to compute the answer when Edison returned and impatiently asked for the results. Upton said he would need more time. "Why," said Edison, "I would simply take that bulb, fill it with a liquid, and measure its volume directly!"

When Upton joined the staff late in October, Edison had already committed himself to the incandescent light. This, rather than the arc light, was the way to imitate the mildness of gas. But the filament glowing in a vacuum had been sought in vain by numerous inventors for half a century. In choosing the incandescent light rather than the arc light he was "putting aside the technical advance that had brought the arc light to the commercial stage." No one, including himself, had succeeded in making an incandescent lamp that would work for more than a few minutes.

Edison's first efforts in 1878 were not notably more successful. Knowing that carbon has the highest melting point of all the elements, he first tried strips of carbonized paper as "burners" and managed to keep them incandescent for "about eight minutes" before they burned up in the partial vacuum of his glass containers. Turning to the infusible metals, he tried spirals of platinum wire; they gave a brilliant light but melted in the heat. Edison accordingly devised a feedback thermostat device that switched off the current when the heat approached the melting point. The lamp now blinked instead of going out entirely. Nonetheless, with his eye on the problem of financing, Edison filed a patent application on October 5 and invited the press in for a demonstration.

As this discouraging work proceeded in the weeks that followed, Edison turned, with Upton's help, to calculating the current that would be consumed by a lighting system equipped with a certain number of such lamps. They assumed that the lights would be connected in parallel, so their imaginary householder could turn one light in the circuit on or off at will, as in a gas-lighting system. Thinking in round numbers, they assumed that these lamps, when perfected, might have a resistance of one ohm and so would consume 10 amperes of current at 10 volts. Allowing in addition for the energy losses in the distribution system, they found that it would require a fabulous amount of copper to light just a few city blocks. Such a system of low-resistance lights was clearly a commercial impossibility.

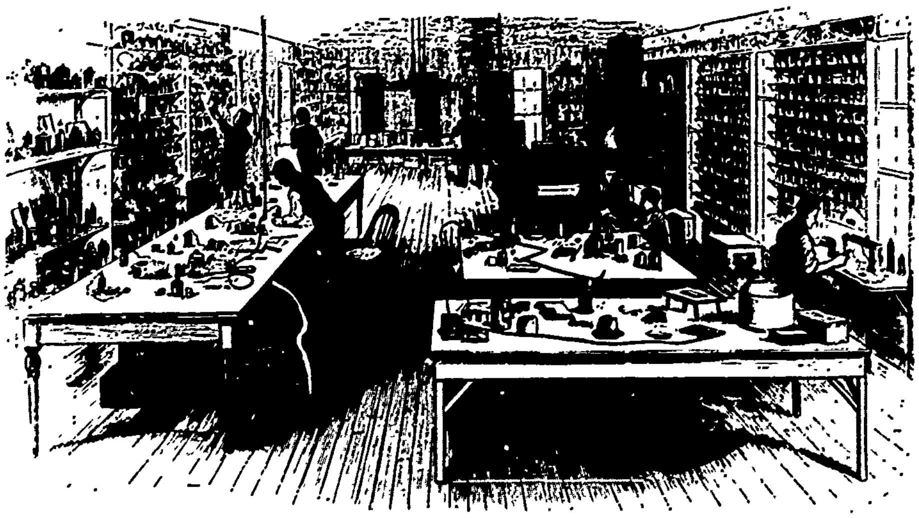
This was the gist of the objections which had greeted Edison's first announcements that he would use an incandescent bulb in a parallel circuit. Typical of the scorn heaped upon him was the opinion expressed by a committee set up by the British Parliament to investigate the crash of gas-lighting securities. With the advice of British scientists, the members of the committee declared that though these plans seemed "good enough for our transatlantic friends," they were "unworthy of the attention of practical or scientific men." From Ohm's law, which governs the relationship between voltage, amperage and resistance in a circuit, the report argued that if an electric light of 1,000 candlepower were divided into 10 smaller lights and connected in parallel, each of the smaller lights would radiate not one tenth but "one hundredth only of the original light." In this judgment such figures as Lord Kelvin and John Tyndall concurred. Before the Royal Institution in London the distinguished electrician Sir William Preece declared: "Subdivision of the electric light is an absolute *ignis fatuus*."

Ohm's law does indeed show that the amount of current (amperes) flowing in a circuit is equal to the electromotive force (volts) divided by the resistance (ohms) in the circuit. Edison's contemporaries reasoned that an increase in the number of lights in a circuit would increase the resistance and therefore reduce the flow of current to each. It was thought that the only way to provide these lights with sufficient current was to reduce the resistance in the distribution system. In a parallel circuit this meant increasing the thickness of the copper conductors to an impractical degree. Such were the limits on the operation of arc lights, with their low resistance and huge appetite for current. Upton's calculations showed that this conclusion also applied to Edison's first low-resistance incandescent lamps.

Edison now confounded his collaborator by proposing that he make the same sort of estimates for an entirely different kind of circuit. This time he would assume lights of very high resistance, supplied with current at high voltage and low amperage. In November and December Upton made calculations on the basis of the same number of lights, but lights with the high resistance of 100 ohms each. These lights were to operate on the low current of only one ampere. Their high resistance was to be offset, in accord with Ohm's law, by the high voltage of 100 volts in the circuit. The result was astonishing: A high-resistance system would require only one hundredth of the weight of copper conductor needed for a low-resistance system. And copper was the most costly element involved—the decisive economic factor.

The High-Resistance System

Here was the crux of Edison's insight at Ansonia. He had recognized there that the subdivision of light called for lamps of high resistance which would consume but little current; to balance the electrical equation it would be neces-

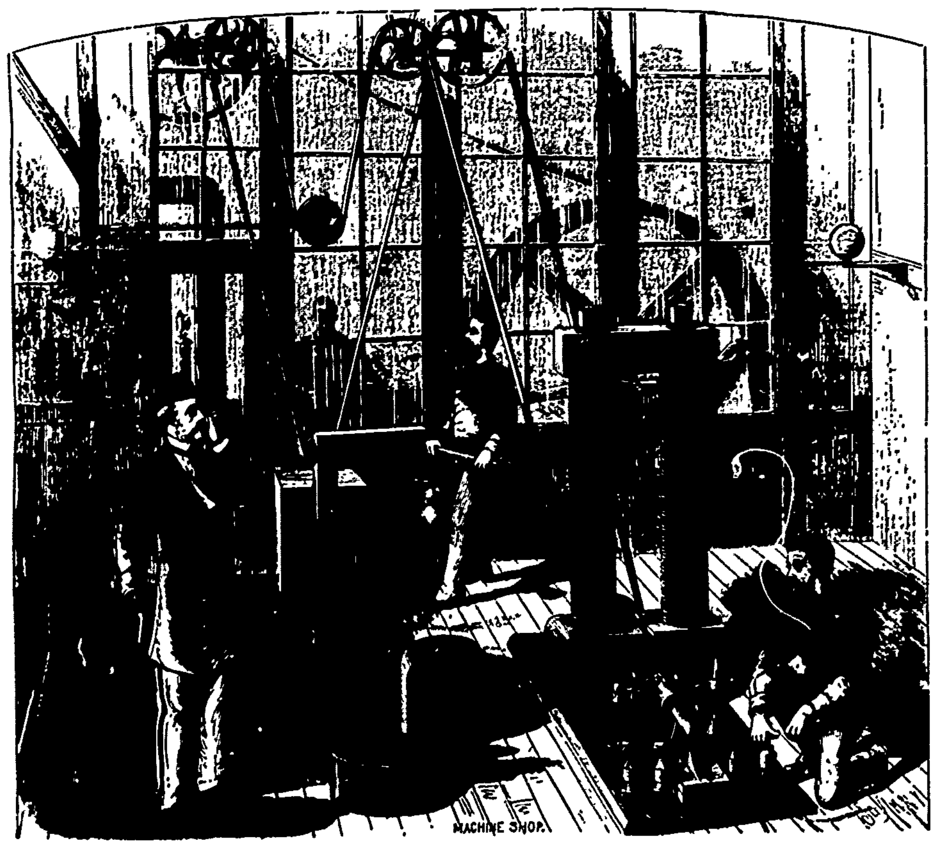


INTERIOR OF EDISON'S LABORATORY at Menlo Park was also depicted in the January 10, 1880, issue of *Frank Leslie's Illustrated Newspaper*. At the time of the work on the electric light the laboratory had expanded into several other buildings.

sary to supply the current at high voltage. This was the "necessary secret" that was "so simple." Today every high-school physics student learns that the power lost in transmitting electric energy varies with the square of the current. Thus a tenfold reduction in current meant a decrease of a hundredfold in the energy wasted (or a hundredfold decrease in the weight of the transmission line). It was a conception easily reached by an elementary application of Ohm's law, but it had not occurred to any of Edison's contemporaries. Even Upton did not immediately grasp the full import of Edison's idea. As he said later: "I cannot imagine why I did not see the elementary facts in 1878 and 1879 more clearly than I did. I came to Mr. Edison a trained man, with a year's experience in Helmholtz's laboratory. . . . a working knowledge of calculus and a mathematical turn of mind. Yet my eyes were blind in comparison with those of today; and . . . I want to say that I had company!"

With Upton's figures before him Edison was convinced that a new and strategic invention lay surely within his grasp. It was clear what kind of distributing system he wanted And he knew what form of incandescent burner would serve his purpose. To offer the necessary resistance to the passage of current it must have a small cross section and so would have a small radiating surface.

By January, 1879, Edison was testing his first high-resistance lamp. It had a spiral of very fine platinum wire set in a globe that contained as high a vacuum as could be achieved with an ordinary air pump. The results were encouraging; these lamps lasted "an hour or two." He then attacked the dual problem of getting a higher vacuum and improving his incandescing element. After another trial with carbon, he returned to metals. platinum, iridium, boron, chromium, molybdenum, osmium—virtually every infusible metal. He thought of tungsten, but could not work it with existing tools. Discouraged by the problem, Edison tried nitrogen in his globe and then resumed his efforts to obtain a higher vacuum. Hearing of the new and efficient Sprengel vacuum pump, which used mercury to trap and expel air, he sent Upton to borrow one from the nearby College of New Jersey (now Princeton University). When Upton returned with the pump late that night, Edison kept him and the other men on the staff up the rest of the night trying it out



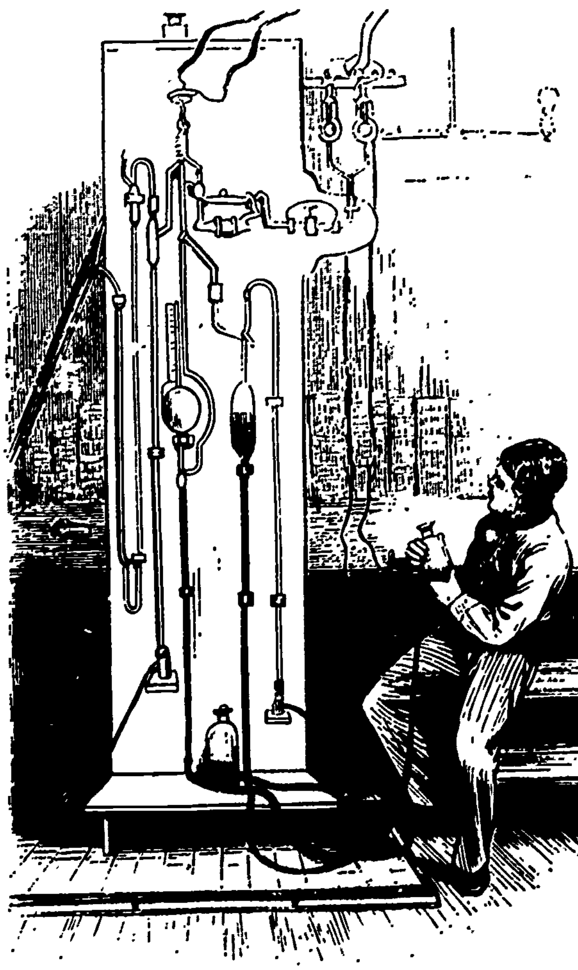
GENERATOR which Edison developed for the needs of electric lighting appears at right in this engraving from SCIENTIFIC AMERICAN for October 18, 1879 (at that time this magazine appeared weekly). The generator was called the "long-waisted Mary Ann."

At this stage Edison made a useful finding: "I have discovered," he noted, "that many metals which have gas within their pores have a lower melting point than when free of such gas." With the aid of the Sprengel pump he devised a method of expelling these occluded gases, by heating the element while the air was being exhausted from the bulb. The platinum wire within the bulb thereupon became extremely hard and could endure far higher temperatures. Edison later said that at this stage he "had made the first real steps toward the modern incandescent lamp."

Meanwhile the spirits of his financial sponsors had begun to droop. Their brilliant inventor, far from having achieved anything tangible, was hinting plainly that he needed more money. The first Brush arc lights were ablaze over lower Broadway, and more were being installed elsewhere with impressive effect. Edison's backers began to have serious doubts as to whether he had pursued the right course To shore up their morale Lowrey arranged to have Edison give them a private demonstration.

In April, as one of Edison's associates recalled it, "They came to Menlo Park on a late afternoon train from New York. It was already dark when they were conducted into the machine shop where we had several platinum lamps installed in series." The "boss" showed his visitors pieces of platinum coil he was using in the lamps, pointed out the arrangement of the lights and described the type of generator he hoped to build. Then, the room having grown quite dark, he told "Honest John" Kruesi to "turn on the juice slowly."

"Today, I can still see those lamps rising to a cherry-red . . . and hear Mr. Edison saying 'A little more juice' and the lamps began to glow. 'A little more.' . . . and then one emits a light like a star, after which there is an eruption and a puff, and the machine shop is in total darkness. . . . The operation was repeated two or three times, with about the same results."



VACUUM PUMP used to remove air from lamp bulbs (top center) was of a new type about which Edison had read in a scientific journal. The man is holding a vessel of mercury.

The platinum coils still consumed a lot of power for the light they gave, and they were costly and short-lived. The temporary Wallace-Farmer dynamos heated up badly, and were not powerful enough to enable Edison to connect his lamps in parallel. Edison admitted that the system was not yet "practical."

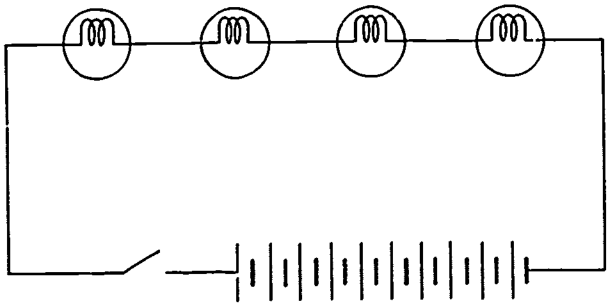
It was a gloomy gathering that broke up on that raw April evening. All of Lowrey's abounding faith would be necessary to rally the spirits and funds of Edison's despondent backers. Some rumors of the disappointing demonstration leaked out; the price of Edison stock fell sharply, while that of gas-lighting securities rose. "After that demonstration," Edison's associate relates, "we had a general house cleaning at the laboratory, and the metallic lamps were stored away."

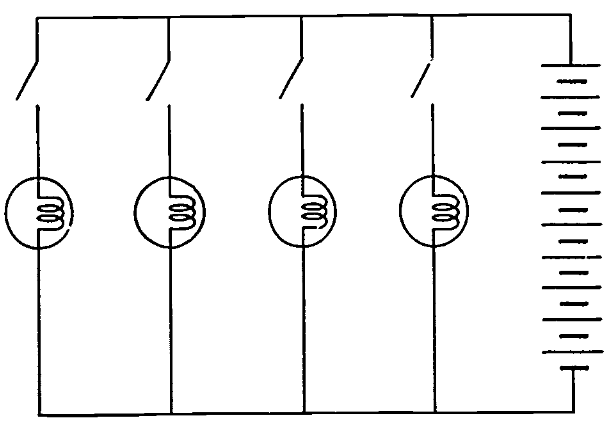
Edison now rallied his staff to efforts on a much broader area of the front "under siege." He followed three main lines of investigation. One group he detailed to the task of developing the dynamo to supply the constant-voltage current required by his high-resistance system. He set another group to pulling down a still higher vacuum in the glass bulbs. The third team, under his watchful eye, carried out the series of experiments in which 1,600 different materials were tested for their worth as incandescent elements.

The "Long-Waisted Mary Ann"

To subdivide the electric current for numerous small lights in parallel Edison needed a dynamo which would produce a higher voltage than any dynamo in existence, and which would maintain that voltage constant under varying demands for current from the system. Existing dynamos were designed around the fallacious notion, held by most electrical experts, that the internal resistance of the dynamo must be equal to the external resistance of the circuit. Through study of battery circuits they had proved that a dynamo could attain a maximum efficiency of only 50 per cent. In 1877 a committee of scientists appointed by the Franklin Institute in Philadelphia had been impressed to discover that the most successful European dynamo, designed by Zénobe Théophile Gramme, converted into electricity 38 to 41 per cent of the mechanical energy supplied to it. The efficiency of the Brush dynamo was even lower: 31 per cent. These machines and their theoretically successful contemporaries all produced current at a relatively low voltage.

Edison had concluded, however, that he must produce a dynamo of reduced internal resistance capable of generating current at a high voltage. Such a machine would not only meet the needs of his lighting system but would also convert mechanical energy to electrical energy with far greater efficiency. As his associate Francis Jehl recalled, Edison said that "he did not intend to build up a system of distribution in which the external resistance would be equal to the internal resistance. He said he was just about going to do the *opposite:* he wanted a large external resistance and a low internal resistance. He said he wanted to sell the energy outside the station and not waste it in the dynamo and the conductors, where it brought no profits." Jehl, who carried out the tests





SERIES CIRCUIT *(top)* requires that a number of electric lights *(circles)* be turned on or off at the same time by a single switch *(break in circuit)*. Parallel circuit *(bottom)*, which was adopted by Edison, makes it possible to turn lights on or off one at a time.

EDISON'S LIGHT.

The Great Inventor's Triumph in Electric Illumination.

A SCRAP OF PAPER.

It Makes a Light, Without Gas or Flame, Cheaper Than Oil.

TRANSFORMED IN THE FURNACE.

Complete Details of the Perfected Carbon Lamp.

FIFTEEN MONTHS OF TOIL.

Story of His Tireless Experiments with Lamps, Burners and Generators.

SUCCESS IN A COTTON THREAD.

The Wizard's Byplay, with Bodily Pain and Gold "Tailings."

HISTORY OF ELECTRIC LIGHTING.

The near approach of the first public exhibition of Edison's long looked for electric light, announced to take place on New Year's Eve at Menlo Park, on which occasion that place will be illuminated with the new light, has revived public interest in the great inventor's work, and throughout the civilized world scientists and people generally are anxiously awaiting the result. From the beginning of his experiments in electric lighting to the present time Mr. Edison has kept his laboratory guardedly closed, and no authoritative account (except that published in the HERALD some months ago relating to his first patent) of any of the important steps of his progress has been made public—a course of procedure the inventor [illegible] absolutely necessary for his own protection. The HERALD is now, however, enabled to present to its readers a full and accurate account of his work from its inception to its completion.

A LIGHTED PAPER.

Edison's electric light, incredible as it may appear, is produced from a little piece of paper—a tiny strip of paper that a breath would blow away. Through

FIRST NEWSPAPER ACCOUNT of Edison's brilliant success appeared in *The New York Herald* for December 21, 1879.

of resistance, also remarked that the art of constructing dynamos was then as mysterious as air navigation. All electrical testing was in the embryonic stage. "There were no instruments for measuring volts and amperes directly: it was like a carpenter without his foot rule."

Upton himself had his difficulties in this hitherto unexplored field: "I remember distinctly when Mr. Edison gave me the problem of placing a motor in circuit, in multiple arc, with a fixed resistance, and I . . . could find no prior solution. There was nothing I could find bearing on the [effect of the] counter-electro-motive force of the armature . . . and the resistance of the armature on the work given out by the armature. It was a wonderful experience to have problems given me by him based on enormous experience in practical work and applying to new lines of progress."

The problem of a constant-voltage dynamo was attacked with the usual Edisonian élan. Seeking to visualize every possible structural innovation for his dynamo armature, he had his men lay out numerous wooden dummies on the floor and wind wire around them, spurring them on in their task by laying wagers as to who would finish first.

After Edison had decided upon the form of winding and type of electromagnets to be used, Upton made drawings and tables from which the real armatures were wound and attached to the commutator. Edison eventually worked out an armature made of thin sheets of iron interleaved with insulating sheets of mica; this armature developed fewer eddy currents and so produced less heat than the solid armature cores then used. When the new cores were test-run, it was Upton who made the mathematical calculations from these tests and drew up the final blueprints.

The self-effacing Upton can be given principal credit for interpreting Edison's ideas and translating them into mathematical form. A careful student of contemporary electrical knowledge, he seems to have been conversant with, and to have guided himself by, the design of a German dynamo, made by the Siemens works, that employed an auxiliary source of current to excite its field magnets.

The new Menlo Park dynamo comprised many admirable features for that period. With its great masses of iron and large, heavy wires, it stood in bold contrast to its contemporary competitors. Owing to the two upright columns of its field electromagnets, it was nicknamed "Edison's long-waisted Mary Ann."

When the dynamo was run at the correct speed, the voltage between its armature brushes was approximately 110, and remained fairly constant, falling but slightly when increasing amounts of current were taken out of the machine. Edison and Upton also contrived a simple but ingenious dynamometer by which the torque of a drive belt was used to measure the work output of the steam engine that powered the dynamo. When Kruesi completed the first operating machine, Upton carefully checked the results. To his astonishment—and quite as Edison had "guessed"—the new dynamo, tested at full load, showed 90-per-cent efficiency in converting steam power into electrical energy.

Edison was as jubilant as a small boy. As was usual with him, the world was soon told all about his "Faradic machine." It was described and depicted in SCIENTIFIC AMERICAN for October 18, 1879, in an article written by Upton.

Once more there was scoffing at Edison's "absurd claims." The hectoring of Edison by some of the leading U. S. electrical experts, among them Henry Morton of the Stevens Institute of Technology, now seems traceable to their ignorance. Reading Morton's predictions of failure, Edison grimly promised that once he had it all running "sure-fire," he would erect at Menlo Park a little statue to his critic which would be eternally illuminated by an Edison lamp.

As a matter of fact, this allegedly ignorant "mechanic" was to be found reading scientific journals and institutional proceedings at all hours of the day and night. It was thus that he had learned about the Sprengel vacuum pump. This device enabled him to achieve an increasingly greater vacuum and to test a broad variety of metals, rare earths and carbon compounds under hitherto unexplored conditions.

The globe itself was also much improved, by the inventor's own design, after he had brought to Menlo Park an artistic German glass blower named Ludwig Boehm. Edison one day drew a sketch of a one-piece, all-glass globe whose joint was completely sealed, and late in April, 1879, Boehm, working skillfully with hand and mouth, fashioned it in the small glass blower's shed in back of the laboratory.

"There never has been a vacuum produced in this country that approached anywhere near the vacuum which is necessary for me," Edison wrote in his notebook. After months of effort he could say exultantly: "We succeeded in making a pump by which we obtained a vacuum of one-millionth part of an atmosphere."

In the late summer of 1879 he realized with growing excitement that a key position had been won. He had a dynamo supplying constant high voltage, and a tight glass globe containing a high vacuum. In his mind's eye he saw what might be done with an extremely fine, highly resistant incandescing substance under these conditions. His state of tension is reflected in the laboratory notebooks by such exclamations as "S . . . ! Glass busted by Boehm!" All that remained for him was to discover a filament that would endure.

The Carbon Filament

In late August or early September—about a year after he first took up his search—he turned back to experimenting with carbon, this time for good. The rods of carbon he had tried earlier had been impossible to handle, as he now understood, because carbon in its porous state has a marked propensity for absorbing gases. But once he had a truly high vacuum and a method for expelling occluded gases he saw that he might achieve better results with carbon than with platinum.

In a shed in back of the laboratory there was a line of kerosene lamps always burning, and a laborer engaged in scraping the lampblack from the glass chimneys to make carbon cake. But lampblack carbon by itself was not durable enough to be made into fine lamp filaments. Edison and Upton had arrived at the conclusion that, given a 100-volt multiple-arc circuit, the resistance of the lamps should be raised to about 200 ohms, this meant that the filament could be no thicker than a 64th of an inch.

Through the summer months Edison and his staff worked at the tantalizing task of making fine reeds of lampblack carbon mixed with tar. His assistants kept kneading away at this putty-like substance for hours. It seemed impossible to make threads out of it; as an assistant complained one day, the stuff crumbled.

"How long did you knead it?" Edison asked.

"More than an hour."

"Well just keep on for a few hours more and it will come out all right."

Before long they were able to make filaments as thin as seven thousandths of an inch. Edison then systematically investigated the relations between the electrical resistance, shape and heat radiation of the filaments. On October 7, 1879, he entered in his notebook a report on 24 hours of work: "A spiral made of burnt lampblack was even better than the Wallace (soft carbon) mixture." This was indeed promising: the threads lasted an hour or two before they burned out. But it was not yet good enough

As he felt himself approaching the goal Edison drove his co-workers harder than ever. They held watches over current tests around the clock, one man getting a few hours' sleep while another remained awake. One of the laboratory assistants invented what was called a "corpse-reviver," a sort of noise machine that would be set going with horrible effect to waken anyone who overslept. Upton said that Edison "could never understand the limitations of the strength of other men because his own mental and physical endurance seemed to be without limit."

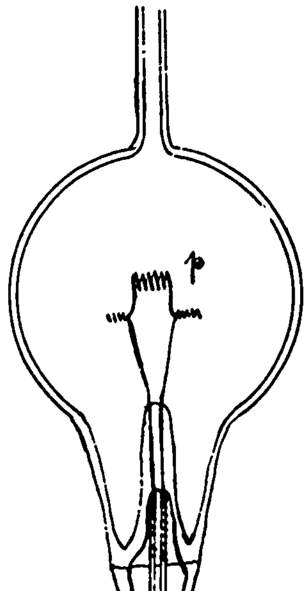
The laboratory notebooks for October, 1879, show Edison's mood of anticipation pervading the whole staff. He pushed on with hundreds of trials of fine filaments, so attenuated that no one could conceive how they could stand up under heat Finally he tried various methods of treating cotton threads, hoping that their fibrous texture might give strength to the filament even after they had been carbonized. Before heating them in the furnace he packed them with powdered carbon in an earthenware crucible sealed with fire clay. After many failures in the effort to clamp the delicate filament to platinum lead-in wires, Edison learned to mold them together with lampblack and then fuse the joint between them in the act of carbonization.

Then, as Edison later related, it was necessary to take the filament to the glass blower's shed in order to seal it within a globe: "With the utmost precaution Batchelor took up the precious carbon, and I marched after him, as if guarding a mighty treasure. To our consternation, just as we reached the glass blower's bench, the wretched carbon broke. We turned back to the main laboratory and set to work again. It was late in the afternoon before we produced another carbon, which was broken by a jeweler's screwdriver falling against it. But we turned back again and before nightfall the carbon was completed and inserted in the lamp. The bulb was exhausted of air and sealed, the current turned on, and the sight we had so long desired to see met our eyes."

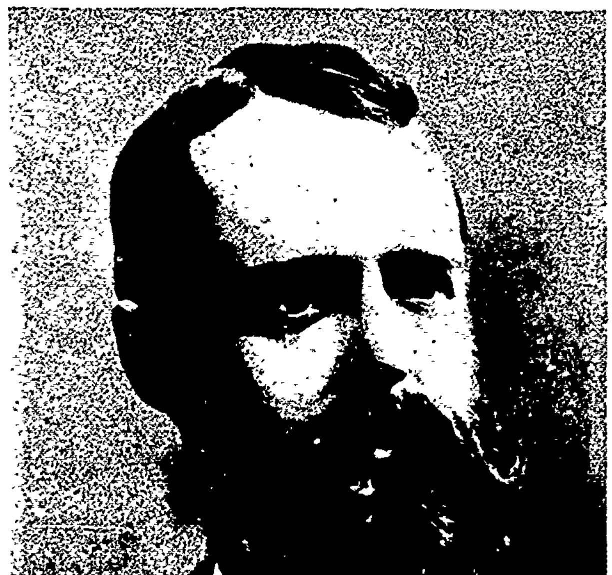
"Ordinary Thread"

The entries in the laboratory notebooks, although bare and impersonal, nonetheless convey the drama and sense of triumphant resolution pervading the laboratory that night. "October 21—No. 9 ordinary thread Coats Co. cord No. 29, came up to one-half candle and was put on 18 cells battery permanently at 1:30 A.M.... No. 9 on from 1:30 A.M. till 3 P.M.—13½ hours and was then raised to 3 gas jets for one hour then cracked glass and busted."

As the light went out the weary men



EARLY EXPERIMENTAL LAMP is depicted in one of Edison's notebooks. This lamp had a filament of platinum. It melted.



FRANCIS R. UPTON made invaluable calculations for Edison's system. An electrical engineer who had studied with Hermann von Helmholtz, he was named "Culture" by Edison.

waiting there jumped from their chairs and shouted with joy. Edison, one of them recalled, remained quiet and then said: "If it can burn that number of hours I know I can make it burn a hundred." Yet all the workers at Menlo Park—Edison, Upton, Kruesi, Boehm and the rest—were completely astonished at their success. They had become accustomed to laboring without hope. "They never dreamed," as one contemporary account put it, "that their long months ... of hard work could be ended thus abruptly, and almost by accident. The suddenness of it takes their breath away."

For once Edison tried to be discreet and keep his momentous discoveries a secret until he could improve upon his lamp filament. At length, after experimenting with various cellulose fibers, he found that paper, in the form of tough Bristol cardboard, proved most enduring when carbonized. Edison was exultant when this filament burned for 170 hours, and swore that he would perfect his lamp so that it would withstand 400 to 1,000 hours of incandescence before any news of it was published.

On November 1, 1879, he executed a patent application for a carbon-filament lamp. Its most significant passage was the declaration: "The object of the invention is to produce electric lamps giving light by incandescence, which lamps shall have high resistance, so as to allow the practical subdivision of the electric light. ... The invention consists in a light-giving body of carbon wire ... to offer great resistance to the passage of the electric current, and at the same time present but a slight surface from which radiation can take place." The specifications called for a distinctive one-piece all-glass container, lead-in wires of platinum that passed through the glass base and were fused to the carbon filament, and joints that were sealed by fusing the glass.

Here were the essential features of the basic Edison carbon-filament lamp, in the form that was to be known to the world during the next half century. It was not the "first" electric light, nor even the first incandescent electric lamp. It was, however, the first practical and economical electric light for universal domestic use.

Edison had spent more than $42,000 on his experiments—far more than he had been advanced by his backers. Now he asked for more money so that he might complete a pilot light-and-power station at Menlo Park. But the directors were still uncertain about the future of the invention. Was it "only a laboratory toy," as one of them charged? Would it not need a good deal of work before it became marketable? Grosvenor Lowrey stoutly defended his protégé. He got no results until he prematurely, and over Edison's objections, made the secret of the electric lamp public.

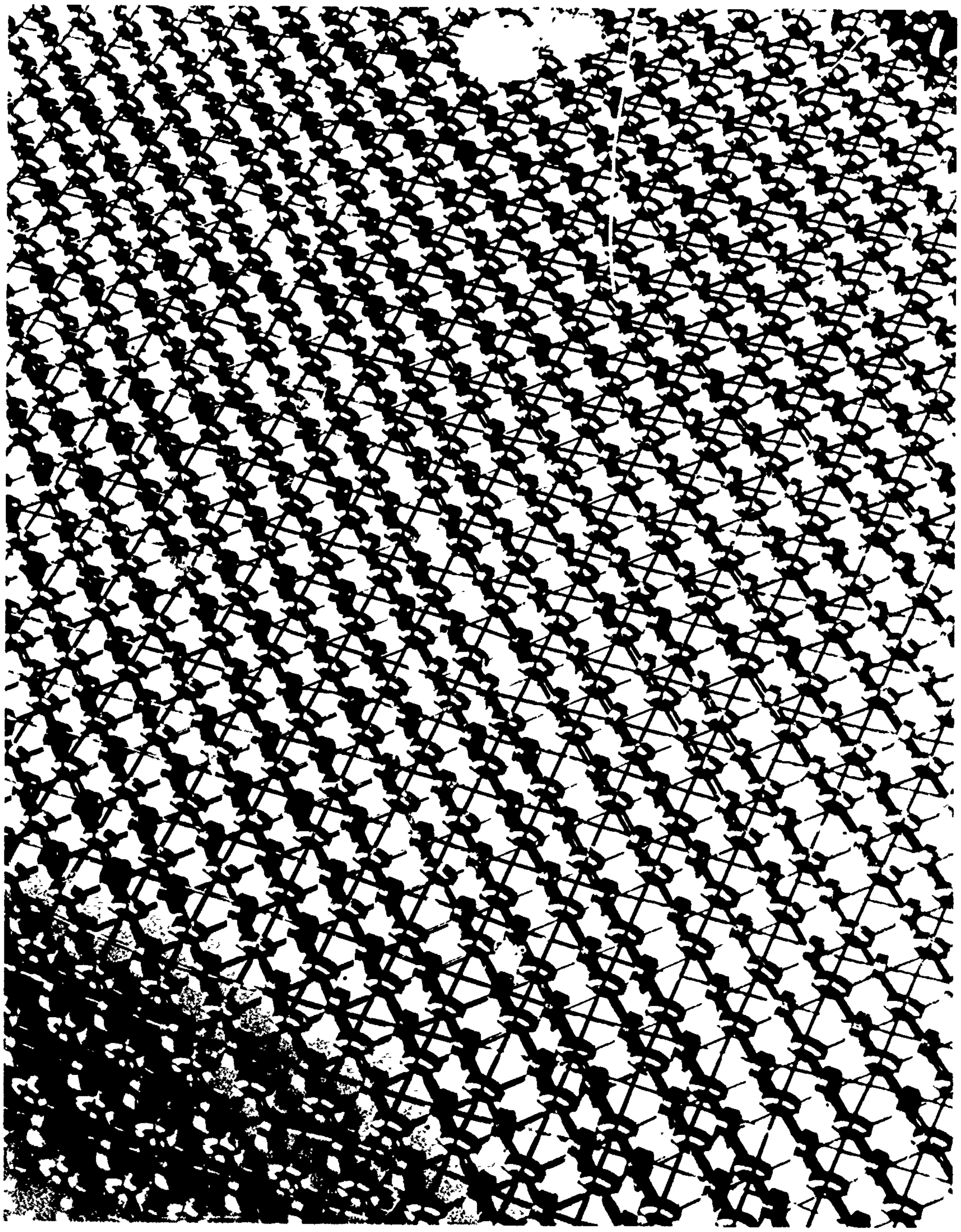
Rumors had been spreading for several weeks. New Jersey neighbors told of brilliant lights blazing all night at Menlo Park, and railroad passengers between New York and Philadelphia also saw the bright lights with astonishment from their train windows. In Wall Street there was a flurry of speculation in Edison stock; the price rose briefly to $3,500 a share.

Then came a front-page story in *The New York Herald* on Sunday, December 21, 1879. There followed an exclusive article about the inventor's struggles for the past 14 months, told to the world, *con amore*, by Marshall Fox, who had written much of Edison before. The detailed treatment of such an adventure in applied science as a feature story was something of an innovation. Also somewhat unusual in the journalism of the time was its relative accuracy of detail, owing to help provided by Upton, who also supplied drawings for the *Herald's* Sunday supplement. The writer did his best to explain how this light was produced from a "tiny strip of paper that a breath would blow away"; why the paper filament did not burn up but became as hard as granite; and how the light-without-flame could be ignited—without a match—when an electric current passed through it, giving a "bright, beautiful light, like the mellow sunset of an Italian autumn."

In the week following Christmas hundreds of visitors made their way to the New Jersey hamlet. Edison hurried with his preparations for an announced New Year's Eve display as best he could, but was forced to use his whole staff of 60 persons to handle the crowds. He could do no more than put on an improvised exhibition, with only one dynamo and a few dozen lights.

The closing nights of the year 1879 turned into a spontaneous festival that reached its climax on New Year's Eve, when a mob of 3,000 sight-seers flooded the place. The visitors never seemed to tire of turning those lights on and off.

The inventor promised the sight-seers that this was but a token of what was in store. He was awaiting the completion of a new generator, he said, and intended to illuminate the surroundings of Menlo Park, for a square mile, with 800 lights. After that he would light up the darkness of the neighboring towns, and even the cities of Newark and New York.



Magnetic core storage element of IBM computer.

9 **High Fidelity**

Edgar Villchur

Two chapters from his book *Reproduction of Sound* published in 1962.

IT MIGHT APPEAR that following a discussion of the nature of sound, the logical subject to consider would be the criteria for reproducing this sound with "high fidelity" to the original. One other element, however, should be covered first—the way in which we hear.

Perception of Sound

We have already seen, in examining units of measurement for pitch and power—the octave and the decibel—that our perception of sound does not necessarily correspond directly to the objective reality. The illusion is consistent, however, so that a given sound always has the same effect on a normal ear.

An important element in the perception of sound was discovered by Fletcher and Munson in 1933. These investigators demonstrated that our impression of loudness did not depend solely on the amplitude of the sound wave, but on other things as well. Specifically, they showed that sound in the lower treble range of the frequency spectrum—the 3500-cps region—appeared to be much louder than sound of the same amplitude at any other part of the spectrum. Thus, if the frequency scale was swept by a tone which continuously rose in frequency but kept exactly the same amplitude, the *loudness*, or apparent amplitude, would increase to a maximum at about 3500 cps and then fall off again.

This fact does not have much practical interest for the person listening to reproduced music, except as it describes the relative nuisance value of different types of noise. No matter how lop-sided our interpretation of acoustic reality, we make the same interpretation in the concert hall as in our living room, and the craftsmen who designed musical instruments (who worked to satisfy their ears, not sound-level meters) perceived sound in the same way.

Fletcher and Munson made a second discovery, however, that does bear directly on the reproduction of sound. They found that the effect described above took place in varying degree, depending on the over-all level of the sound. For very high amplitude sound the drop in loudness with frequency below 3500 cps hardly occurred at all, while for very soft sound the effect was maximum. Above 3500 cps the effect remained constant, within 2 or 3 db, no matter what the over-all sound level.

The well-known "equal loudness contours," also referred to as the Fletcher-Munson curves, are reproduced in *Fig.* 2–1. Each curve plots the sound amplitude required to produce the same perceived loudness at different frequencies of the scale. It can be seen that normal hearing losses in the bass end become progressively greater as the over-all sound level is decreased.

This means that if an orchestra plays a musical passage at the sound level represented by 90 db, and if this music is reproduced at the 60 db level, we will hear the bass with less *relative* loudness than we would have heard it at the concert itself. If you follow the 90- and 60-db curves, shown superimposed in *Fig.* 2–2, you will see that there is approximately a 14 db perceived loss at 50 cps—it takes 14 db more of actual amplitude, in the lower curve, to produce the same relative loudness at 50 cps as it does in the upper curve.

In order to re-create the original balance of perceived frequencies at low volume levels, it has become customary to introduce bass boost which is related to the setting of the volume control, either automatically or otherwise.

A volume control tied to automatic bass boost is called a *loudness control.* (Some loudness controls also boost the treble spectrum appreciably at low volume settings. There is no justification for this in the Fletcher-Munson curves.)

High Fidelity to What?

The assumption will be made here that the purpose of high fidelity equipment is to reproduce as closely as possible the experience of the concert hall, not to transcend or improve it.

I remember an exhibition at New York's Museum of Modern Art, during the late thirties, of "high fidelity" reproductions of water color paintings. Life-size reproductions were hung side by side with the originals, and it was often difficult or impossible to tell them apart. There was no question in anyone's mind about how to judge the quality of these prints. The only criterion was accuracy. The public that visited the exhibit was used to looking at paintings, and was able to make an immediate comparison

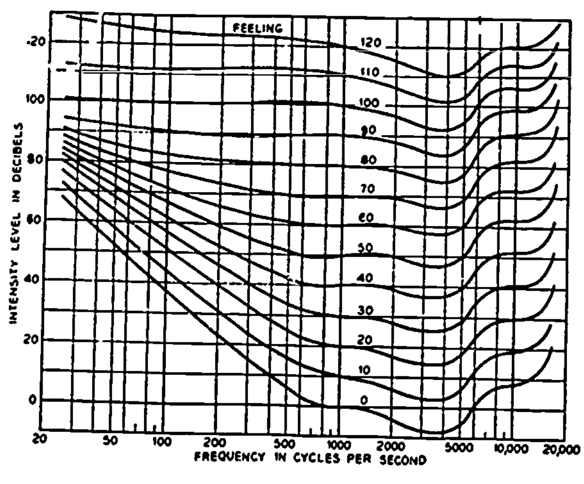


Fig. 2-1. The Fletcher-Munson equal loudness contours. For each curve, the height at any point represents the sound amplitude required to produce the same subjective loudness as at 1000 cps. (After Fletcher and Munson)

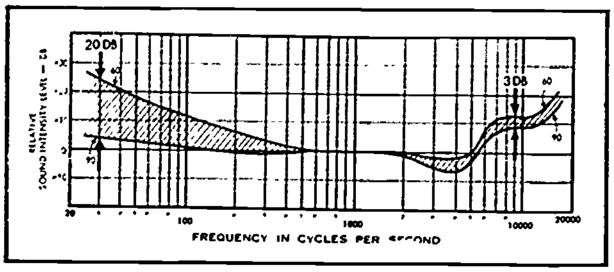


Fig. 2-2. The 60 and 90 db Fletcher-Munson curves superimposed. The shaded area represents the difference in normal hearing loss from one sound level to the other.

between the copy and the original. No one thought of the prints as entities in themselves, with qualities independent of the qualities of the originals.

This point of view does not always hold in the field of high fidelity musical reproduction. Only a minority of today's high fidelity public are concert-goers. Many have never attended a live concert; they know the sound of the orchestra or of individual musical instruments only as it is reported by amplifiers and loudspeakers. They may know what they like in reproduced sound, but they have no way of evaluating the realism of reproduction.

This partly explains why so much variation is tolerated in audio equipment. The same record may sound very different when played through different brands of equipment, each brand equally acceptable in the market place. The evaluation of high fidelity components is popularly thought of as an entirely subjective matter, like comparing the tone of one violin to that of another rather than like holding a facsimile up to its original.

For similar reasons high fidelity demonstrations such as the annual Hi-Fi shows can get away with a lot of sound that is startling but essentially non-musical. Some of the "reproduced" sound that greets the show visitor is necessarily unfamiliar because it has no live counterpart. A harmonica blown up in volume to the dimensions of a theatre organ is a new and different instrument. A crooner whispering into a microphone an inch away invents a new sound; his unamplified voice is never heard in public. A combination of Bongo drum, chimes and electric guitar creates a *tutti* which one may like or dislike, but for which there is no equivalent in one's memory to serve as a live standard.

Such sound can only be accepted as a self-sufficient entity, like an old calendar chromo. Any resemblance to live music or to painting is purely coincidental, and the science and/or art of reproduction is not really involved.

High fidelity has undoubtedly increased rather than decreased the ranks of music lovers, and there are probably more people than ever who are unimpressed with gimmick sound. Many designers and manufacturers in the field work only for naturalness of reproduction. The designer of integrity avoids like the plague those exaggerations that sometimes attract the novice—over-emphasized bass for "depth," over-emphasized mid-range for "presence," over-emphasized treble for "brilliance." These distortions are more properly called, respectively, boominess, nasality or "honkiness," and harshness.

Many demonstrations are not, fortunately, of the gimmick type, and use musical material played at musical levels. There have also been concerts staged with live musicians, in which direct comparisons of reproduced sound to the sound of the live instruments could be made, in the same way that direct comparisons of prints to original paintings were made at the Museum of Modern Art. The live vs. recorded public concert is one method of giving direction to equipment designers and perspective to high fidelity consumers. Although transferring concert hall atmosphere to the home has special problems of its own, success in creating an identity of sound in the concert hall itself solves the major part of the problem. Even more vital to maintaining balance and perspective in the high fidelity world is live concert attendance.

We are now prepared to discuss the technical standards of quality that may be applied to a sound reproducing system. There will be no dividing lines proposed, at which low fidelity becomes medium, high, or super.

Frequency Response

The frequency response of a sound reproducing system, or of one of its components, describes its relative handling of parts of the input signal which differ in frequency. "Handling" may refer to electrical amplification, as in an amplifier, to conversion of mechanical to electrical energy, as in a pickup, or to conversion from electrical to acoustical energy, as in a loudspeaker.

There are two aspects of frequency response: the *range* of frequencies handled, and the *uniformity* with which the unit or system responds to different frequencies. Knowledge of the first of these is useless without knowledge of the second. Let us therefore pass over the question of range for the moment, and determine what uniformity will be required for the range we finally decide on.

Uniformity of Response

Although the trained ear can usually perceive a change of sound level of a db or less in test signals, the average observer is probably less sensitive to a change of sound level in a particular frequency range of a musical passage. Reproduction which remains constant over its frequency range within one or two db would thus probably be adequate for perfect apparent fidelity, other things being equal.

This standard can be met in amplifiers without much difficulty, even at high power levels. The best pickups are also able to conform, but loudspeakers are laggard in this respect.

The results of non-uniform reproduction are several. Undue volume in a particular section of the sound spectrum can produce stridency or boominess as opposed to natural musical sound. More particularly, the existence of sharp peaks in the response curve, usually representing a resonant condition, mean that *hangover* or *ringing* may be present—the speaker cone or section of cone will continue to vibrate after the signal has stopped. This is perceived as a "rain-barrel" effect, a muddying up of the sound and impairment of the distinctness of the different instrumental voices. Such an effect is also indicated when the listener is unable to distinguish clearly the pitch of low-frequency tones.

Another important effect of peaked frequency response is the exaggeration of unwanted noise components such as turntable rumble or record surface scratch. This effect was not given its due recognition in the earlier days of high fidelity, when the existence of rumble and surface noise was proudly displayed as evidence of extended frequency range.

The amount of surface noise in a good quality modern LP record and the amount of rumble from a good record player are such that there will not be much significant noise produced in a system with uniform frequency response, even though the frequency range be extended to the limits of the present state of the art. In a comparison test conducted recently between two tweeters, the one which was able to reproduce almost an octave more of treble (into the inaudible region) showed a dramatic decrease of surface noise, due to its extreme evenness of response. There was no selective reproduction of discrete frequency regions, and the switch to the superior speaker produced a fuller, more natural treble simultaneously with the reduction in surface noise.

A similar situation exists with regard to turntable rumble. A peaked system whose response falls off rapidly below 60 cps may exhibit more turntable rumble than a smooth system whose full response extends an octave lower.

Tell-tale evidence of the existence of peaked reproduction in the bass may be gathered from listening to the reproduction of speech. The male speaking voice ordinarily contains no sound components whose frequency is below 100 cps, and the reproducing system should give no hint (by a boomy, resonant quality in the voice) that it is also capable of speaking in the tones of the double bass.

Range of Response

It is generally agreed among acoustics authorities that the range of 40 to 15,000 cps is sufficient for perfect or near-perfect apparent fidelity in the reproduction of orchestral music. The phrase "near-perfect" is meant to imply that when such a range has been achieved the designer should direct his attention to inaccuracies of reproduction more gross than are associated with the frequency limitations indicated.

For the pipe organ enthusiast, however, there is significant intelligence (significant, that is, from the point of view of the emotional impact of the music) down to 32 cps or lower. 32.7 cps is three octaves below middle C relative to A-440, and is the lowest note of the average pipe organ, although many larger organs reach down an octave lower. These low organ tones are distinguished by the fact that they contain a strong fundamental component. The lowest tones of the piano, on the other hand, contain no fundamental energy that significantly affects the quality of the sound. Even though the lowest key on the piano strikes 27.5 cps, response down to this frequency is not required for the reproduction of piano music.

Probably no characteristic of audio components is so freely booted about by advertising copywriters as frequency range. Any numerical range of frequencies listed is totally meaningless unless accompanied by a description of the decibel tolerance above or below reference that is being used, and, for a loudspeaker, by a description of off-axis response as well. A 3-in. speaker made for portable radios will "respond" when stimulated by a 30-cps signal—perhaps by having its cone tear loose and fly out into the air—and almost any speaker, even a woofer, will make some kind of sound when stimulated by a high-powered 15,000-cps signal. A frequency response rating must mean something more than that a signal of given frequency makes a speaker move audibly, or that it makes an amplifier show an electrical output of some sort at its terminals. It must mean that within a stated frequency range, and, for power devices, within a stated range of power, the fundamental output of a given device is uniform to a stated degree.

Treble Dispersion

The on-axis response of a loudspeaker may be very deceiving, because the higher frequencies tend to be directed in a beam which continually narrows as the frequency is raised. Good sound dispersion must therefore be a qualifying factor for any treble response curve.

A speaker which has relatively uniform treble output both on-axis and off-axis (over a reasonably large solid angle—perhaps 45 degrees in any direction from the axis) will reproduce music with a "spaciousness" that does not exist when there is more concentrated beaming of the treble. Furthermore, severely attenuated off-axis response in the treble means that the total sound power radiated at treble frequencies is considerably less than that implied by the on-axis response curve. It is this total radiated power, rather than the on-axis pressure, that determines whether a speaker will sound dull, natural, or over-bright in a normally reverberant room.

Transient Response

Transient response refers to the accuracy of reproduction of the wave envelope, and is concerned with the reproduction of attack and decay characteristics of the sound. We have seen that uniform frequency response predicts the absence of ringing; if the steady-state frequency response curve does not have peaks, the reproduced sound will die away just as in the original.

Consider, for example, the tone represented in (A) of *Fig.* 2–3. Perfect reproduction would produce an identical wave form, differing perhaps only in amplitude, while poor transient response would be indicated by the hangover that is apparent in (B). The continuation of the reproduced signal after the original has ended may be compared to a color smear on a reproduced painting.

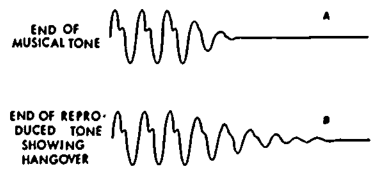


Fig. 2-3. Poor transient response.

Attack time involves the reproduction of frequencies higher than the fundamental. Although a percussive tone may have a low fundamental pitch, the frequency components associated with its steep attack characteristic may be very high. Natural reproduction of a drum beat through a two-way speaker system may thus be accomplished by the "woofer" handling the fundamental tone and its proper decay, while the "tweeter" contributes the sound components that make up the sharp attack.

Harmonic and Intermodulation Distortion

Reproducing devices have a characteristic way of performing with less than perfect accuracy. In addition to the frequencies at which they are asked to vibrate mechanically (or alternate electrically) they introduce new modes of oscillation of their own—and these new frequencies are harmonics, integral multiples of the original frequency. This inaccuracy is called *harmonic distortion*. It is measured as the ratio of the amplitude of the spurious harmonics to the true signal, in per cent.

We have seen that harmonics of fundamental frequencies are produced in any case by musical instruments. Yet small amounts of harmonic distortion produce very unpleasant effects. The sound becomes harsh, unmusical; the bass is wooden and the treble painful.

The primary reason for this is that with harmonic distortion comes an attendant evil—intermodulation distortion. Intermodulation distortion can be described as the introduction of new sound components, at sum and difference frequencies, when tones of two or more frequencies are passed through a non-linear system—that is, a system which creates harmonic distortion. These sum and difference frequencies are harmonically unrelated to the original musical tones. They are musically discordant, and they serve to create raucous, unmusical sound in a degree proportional to their relative strength. The formation of intermodulation products is illustrated in *Fig.* 2–4.

The primary importance of low distortion has always been recognized by audio authorities. It has also become in-

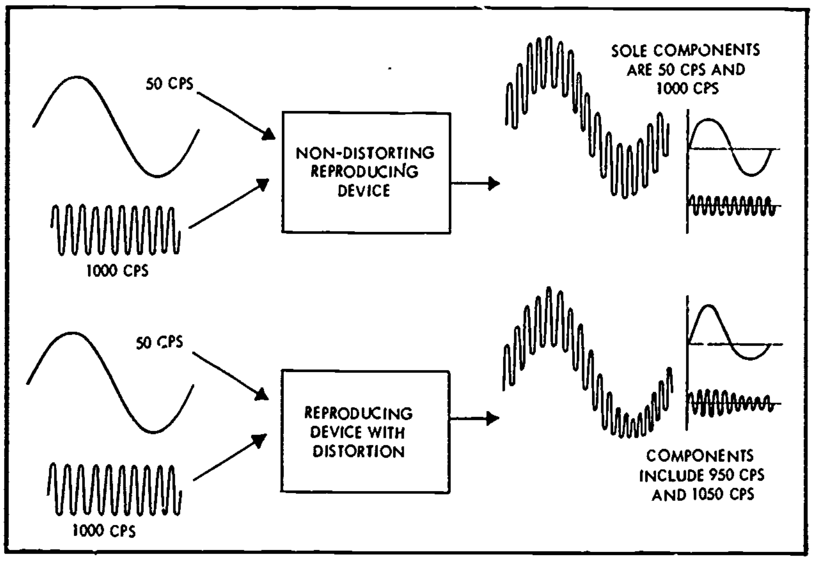


Fig. 2-4. Intermodulation distortion as a result of harmonic distortion of the low-frequency wave form. Note that the wave envelope of the high-frequency tone is "modulated."

creasingly recognized by the high fidelity public in recent years, after the first flush of excitement over reproducing regions of the frequency spectrum previously untouched. Amplifier manufacturers now feature distortion data over frequency response data; unfortunately it is very rare for loudspeaker specifications to make any quantitative reference to distortion at all. The reason lies in the fact that while both harmonic distortion and intermodulation distortion (the latter is usually greater by a factor of 3 or 4) can be kept to extremely low values in high quality amplifiers—a small fraction of one per cent at rated power—the corresponding values for loudspeakers are much higher. In the octave below 60 cps it is a rare speaker indeed which can hold harmonic distortion, at any appreciable sound level, below the 5 per cent mark over the entire octave, and many speakers produce percentages of distortion in this frequency region ten times as great. But the listening results are not as bad as might appear at first glance: speaker response is normally severely attenuated in this lower range, which helps, and there is comparatively little musical material of such low frequency to be distorted.

When the reproducing system has a minimum of low frequency distortion, very low bass tones of high power, such as might be produced by organ pedal pipes, not only remain pure in timbre themselves but do not create intermodulation with the rest of the music; they do not destroy the purity of the treble by introducing false tones.

Power Capability

The power capability of a high-quality reproducing system should be such as to be able to establish an intensity level of sound in the living room equal to the level at a good seat in the original concert hall. The electrical power required of the amplifier for achieving this goal depends upon the efficiency of the speaker, and the sound power required of the speaker depends on the size and other acoustical characteristics of the room. Concert-hall level can be established in a living room with a tiny fraction of the acoustical power of a symphony orchestra, because the lower power is concentrated in a much smaller area.

"Concert-hall level" is sometimes misinterpreted to mean the sound level which would be created if the orchestra were somehow jammed into the living room itself. The writer has yet to experience at a live concert, even during *fortissimo* passages, an assault on his ears that compares to hi-fi assaults he has weathered. It is interesting to note that certain hi-fi demonstrations preclude intelligible conversation which is not shouted, while whispered conversations in a concert hall are liable to prove extremely distracting and annoying to one's neighbors. It is the sound intensity level at the ear, not the power of the orchestra, that we are trying to reproduce.

Noise Level

Any sound component not present in the original program material, other than distortion products, is referred to as noise, even though it may be periodic and not conform to our strictly scientific definition. Hum, rumble, surface scratch, tube hiss or other circuit noise and similar disturbances tend to destroy the auditory illusion, and must be kept to a minimum.

A standard for satisfactorily low noise has been established by the FCC for FM broadcast stations. It is that the power ratio of the maximum signal to the noise must always be at least 60 db; this represents a ratio of one million to one.

Dynamic Range

The dynamic range, or range of amplitude of the reproduced sound from softest to loudest, is determined by the two factors just discussed, noise level and power capability.

Soft musical passages can be masked by any of the types of noise referred to, and therefore the lowest sound levels that can be used must be much louder than the noise level. The maximum sound levels that can be used, of course, are limited by the power capability of the system.

A dynamic range of 60 db, or a million to one power ratio between highest and lowest sound levels, is generally considered adequate for reproduction of the largest symphony orchestra.

Stereo

All of the above considerations apply equally to monaural and to stereophonic reproducing systems. These objective elements of equipment fidelity—low distortion, adequate frequency response, dynamic range, etc.—are able, in stereo, to contribute more to the subjective illusion of musical reality than in a monaural system.

A stereo record-reproduce system has in effect two parallel and complete monaural systems. The work of each component along the way is done twice. The sound is picked up by two separate microphones; the output of each microphone is recorded on a separate track of the tape; the record groove, although not doubled, is cut in such a way as to independently contain the record of each signal channel; the pickup contains tw separate generating elements which independently sense and transmit each signal channel; the two signal outputs of the pickup are sent through independent amplifiers and fed to two independent loudspeakers. There are variations on this ideal scheme, but the above describes the basic concept of stereo.

The purpose of this dual-channel reproduction is, in the simplest terms, to help recreate the acoustical atmosphere of the concert hall. In the old-fashioned stereopticon each visual channel gave a slightly different perspective view of the subject. Similarly, in stereo recording, each microphone gets a slightly different auditory perspective. It is important to note that this auditory perspective is of the orchestra or soloists *in the hall* in which they are performing, not merely of the musical performers in the abstract. This is important because a good part of the sound that reaches our ears at a concert does not come directly from the orchestra, but is reflected from the walls and ceiling of the concert hall.

The channels of a stereo system are identified as "right and "left." This does not mean that one microphone picks up the sound of the right section of the orchestra only, and that the other microphone picks up the sound from the left section of the orchestra. It does mean that one microphone has a right-oriented perspective of the total sound in the recording hall, and that the other microphone has a left-oriented perspective of the total sound. When these two recorded channels (which, like the two photos on a stereopticon card, are very similar to each other) are reproduced through two separate loudspeakers they create, although not perfectly, the illusion of the acoustical environment and sense of space of the concert hall. There is an increased awareness of the physical position of different instruments, but this is very much less important than the general increase in realism and the consequent increase of clarity, particularly from the point of view of the distinctness of the different musical voices.

There is an approach to stereo recording, commonly referred to as "ping-pong" stereo, which provides an exaggerated separation between the right and left channels. If only the left side of the orchestra were playing during a particular passage, there would be practically no sound from the right recording channel. The left-right orientation of the different instruments is the primary goal in this case, rather than reproduction of the original acoustical environment. The degree to which one's attention is directed to the physical position of the instruments in "ping-pong" stereo is often much greater than that at the live concert itself.

The greatest benefit of good stereo recording and reproduction is that it frees us, to a greater extent than was possible previously, from the acoustical environment of the listening room, and transports us to some extent to the acoustical environment of the hall in which the recording was made. The normal living room does not provide the proper acoustical atmosphere for a musical concert, particularly of a large orchestra. Musical instrument designers worked in terms of the tonal qualities that would be produced in the type of concert hall with which they were familiar.

THE SOUND REPRODUCING SYSTEM

THE PHONOGRAPH is a classic example of an invention that cannot be credited wholly to one man. In 1877 Edison directed his assistant, John Kruesi, to construct the first complete record-reproduce system, but sound recorders were sold on a commercial basis as early as 1860, and Thomas Young's "A Course of Lectures on Natural Philosophy" described and illustrated a crude but practical sound recorder in 1807.

Young's recorder consisted of a sharp metal stylus held by spring tension against a revolving cylinder, the cylinder coated with wax and turned by a governor-controlled gravity motor. When a vibrating body such as a tuning fork was held against the stylus, a wavy line was cut into the wax. This line represented the wave form of the vibrations, and it could be studied and analyzed at leisure. The recorder was a mechanical draftsman, that could sense very small motions and record pressure changes that took place within a period of a very small fraction of a second.

By 1856 Léon Scott de Martinville had constructed the "phonautograph" (self-writer of sound) illustrated in *Fig.* 3–1. The sound wave form was scratched by a hog-bristle stylus on the surface of a cylinder coated with lampblack, but the big advance over Young's machine was the fact that the phonautograph could record directly from the air. The force of the acoustical vibrations

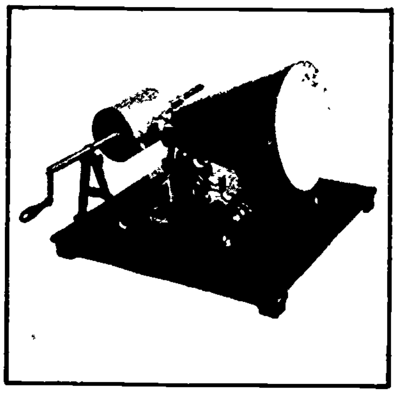


Fig. 3-1. The phonautograph of Léon Scott de Martinville — a commercial sound recorder of the eighteen sixties. (Courtesy Smithsonian Institution)

was concentrated by a horn onto a diaphragm, and the stylus was attached to the diaphragm, so that the recording needle did not have to actually touch the vibrating source of sound. This device, which corresponds in function to the modern oscilloscope, was a catalogue item of the Paris firm of Koenig, and was sold as a measuring instrument to acoustical laboratories.

The phonautograph which is at the Smithsonian Institution at Washington would undoubtedly reproduce music if a proper record were placed on its revolving cylinder. The theoretical possibility of playback was understood then, too, but the lampblack records were useless for playback, as their grooves were not rigid enough to direct the vibrations of a playback needle. About half a year before Edison got his brainstorm Charles Cros conceived a method for bringing the groove sinuosities back to life as sound. The lampblack recording was to be photo-engraved on a metal cylinder, and running a needle through the hard groove would then cause the needle to vibrate from side to side, in the same time pattern as the hog bristle stylus that first inscribed the line.

For reasons which may be related to nineteenth century differences in tradition between the scholar and the industrial engineer, Cros didn't even construct a working model, but merely filed a complete, sealed description of his system with the *Académie des Sciences.* On the other hand, less than a month after Edison first conceived of a reproducing phonograph the country was reading about a working unit in newspaper headlines. There was a great stir of excitement over this amazing tonal imitator, (see *Fig.* 3–2) with public demonstrations, lectures before august scientific bodies, and a visit to the White House.

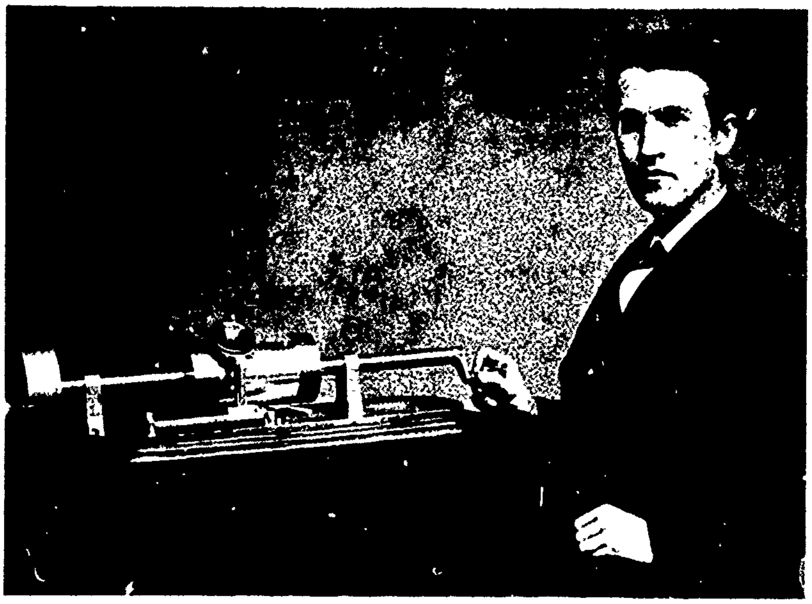


Fig. 3-2. Edison with his tin-foil phonograph. (Photograph by Brady — courtesy Smithsonian Institution)

The excitement soon died down, as the Edison machine was an impractical toy, with neither permanent records nor usable fidelity. The recorded groove was indented into a semi-hard material, tin foil; it was only able to retain its shape partially, and that for very few playings. Subsequent technical improvements, however, made the phonograph a popular device by the turn of the century. It is curious that our modern recording system, in which the record is a mechanical copy of the original master, is more closely related to Cros' system than to Edison's. Emil Berliner, the father of the moulded or cast record, began his research work by successfully carrying out Cros' proposals.

The Mechanical or "Acoustic" Phonograph

It would be useful to consider the design of the non-electric phonograph, as illustrated in (A) of *Fig.* 3–3. A better insight can thereby be gained into the function of the various components of a modern electronic system.

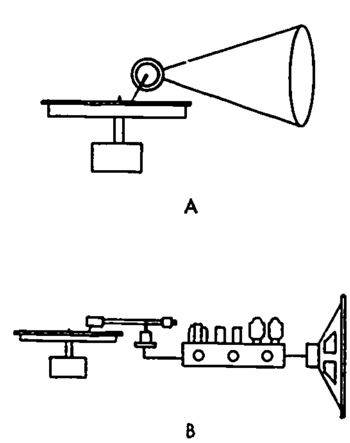


Fig. 3-3. (A) The mechanical phonograph. (B) The electric phonograph.

The wave forms frozen into the record groove control the vibrations of the playback stylus when the groove is dragged past the stylus by a revolving turntable. These stylus vibrations, although they contain a fairly large amount of mechanical energy, engage practically no air, like the revolutions of a bladeless electric fan. The needle is therefore attached to a diaphragm, which vibrates in sympathy with the stylus and has a much larger surface area in contact with the air of the room.

But even the reproducing diaphragm doesn't get a sufficient bite of the air for practical purposes. Therefore the diaphragm is placed at the narrow throat of an acoustical horn, and the actual usable sound emerges into the room from the much larger mouth of the horn. The system works somewhat as though the diaphragm area were really that of the horn's mouth.

It can be seen that all of the energy radiated by the horn is taken from the mechanical vibrations of the needle, and the forces between needle and record groove are necessarily great. This has obvious implications for record wear, but perhaps more important, the demands for power placed on the "sound box" or "speaker" (old-fashioned terms for the needle-diaphragm-head assembly) place a severe limitation on musical fidelity. High distortion and peaked and severely limited frequency response are to be expected.

The Phonograph Amplifier

The solution to this problem lies in changing the function of the phonograph pickup, from the primary generator of sound power to a device which controls an outside source of power. If the power from the outside source is made to oscillate in imitation of the needle vibrations, two benefits can result:

1. The final output sound derived

from the record groove can be much louder.

2. The power demands on the pickup itself are no longer heavy. The pickup can be designed for quality rather than loudness; the problems of achieving uniform, extended frequency response and low distortion are considerably lessened. So, incidentally, is the required weight on the pickup and the grinding away of the record groove.

The control of an outside source of power to conform to given oscillations is called *amplification*. The first phonograph amplifier was pneumatic: the needle was made to actuate an air valve, which periodically throttled a flow of compressed air. Most of the work of radiating sound power was thus performed by the air compressor, and the stylus was relieved of part of its burden.

All modern sound reproducing systems use amplifiers, but unlike the first pneumatic systems these amplifiers are electronic. The phonograph pickup is no longer a sound generator but an electric generator. It produces small alternating voltages at its terminals, whose wave forms conform to those of the groove and of the recorded sound. The pickup has to generate very little power, because the output voltage can be amplified to almost any desired degree. The amplified electrical power must finally, of course, be converted back into sound bv a loudspeaker. The two types of reproducing system, electrical and purely mechanical, are shown in *Fig.* 3–3.

The Modern Sound Reproducing System

The purpose of the historical approach used above has been to furnish the reader with an appreciation of the reason for the modern audio system being designed as it is. With the electronic amplifier supplying the brute force, so to speak, the mechanical components—pickup and loudspeaker—can be built in such a way as to suppress the natural resonant tendencies inherent in mechanical vibratory systems

Before discussing each of the audio components in detail, it would be useful to make a brief survey of the entire reproducing system. A complete monaural system is illustrated in *Fig.* 3–4.

First of all the disc record must be revolved by a *motor* and *turntable*. The chief operational requirements of this part of the system are that it revolve at the correct speed, that the speed be constant, and that extraneous vibrations do not communicate themselves to the pickup.

The first of these requirements is for the purpose of keeping the reproduced music at the same absolute pitch at which it was recorded: too fast a turntable speed will make the pitch sharp, and too low a speed will make it flat. The second condition listed, constant speed, is required in order to avoid pitch variations, or "wow." The third requirement, lack of extraneous vibrations, keeps low-frequency noise called "rumble" out of the final sound.

The groove variations are sensed by the *needle*, or *stylus*, which in high-quality systems is jewel tipped; it is usually diamond. The needle must have an unmarred, smooth surfaced, hard tip, normally of spherical shape.

The *pickup* is an electric generator (usually either of the piezo-electric, variable reluctance, or moving-coil type) whose function is to translate the mechanical vibrations of the needle into electrical oscillations of the same wave form. It must do this with minimum distortion of the wave form, and must not allow resonances of its own to influence its output voltage significantly. It is also an advantage for the pickup to impose as little work as possible on the needle. The greater the force required for the groove to displace the needle from side to side, the greater the vertical bearing force will have to be to

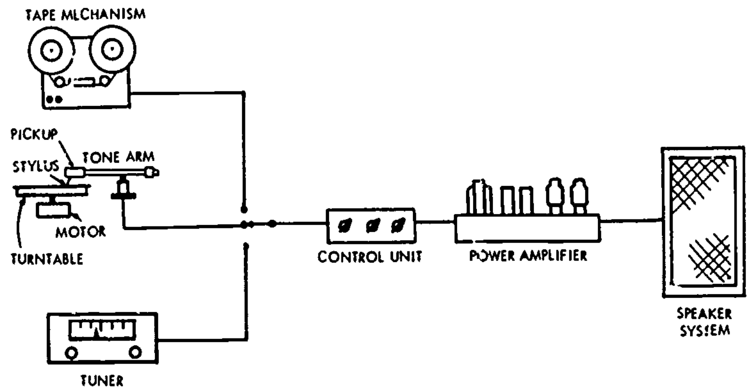


Fig. 3-4. Diagram of a complete monaural sound reproducing system.

maintain proper and constant stylus-groove contact, and the greater the wear of both record and needle.

The *tone arm* holds the pickup in place over the groove, and must provide sufficient freedom of motion so that the pressure of the groove walls alone can make the needle move across the record, following the recorded spiral. It must also be free enough to follow warp and eccentricity of the disc easily. The tone arm must hold the pickup approximately tangent to the groove being played, must provide the proper vertical force for the pickup, and must not allow its own resonant behavior to influence the system.

The electrical output of one type of pickup, the piezo-electric, is usually fed directly to the amplifier. It is of the order of ½ volt or more, and is a fairly accurate replica of the recorded sound. This is so because the characteristic frequency response of the pickup is more or less the inverse image of the frequency characteristics "built in" to the record. (This last subject will be taken up in detail later.)

The reluctance and moving-coil pickups, however, produce a much smaller amount of electrical energy. The output voltage of these pickups (which are classed together as *magnetic* types) may be as low as a few thousandths of a volt. Furthermore the characteristic frequency response of the magnetic pickup does not compensate for the way in which the frequency characteristics of the recorded sound has been doctored. Therefore the pickup output must be passed through a *preamplifier* before it enters the amplifier proper.

The preamplifier is normally combined with the main amplifier control sections (volume and tone controls). Its functions are to increase the output voltage of the pickup, and to compensate accurately for the frequency characteristics of the record so that the sound is not deficient in bass and heavy in the treble. Since different record companies have made records with different characteristics the preamplifier may allow the operator to choose between several types of frequency compensation. The need for such control, which is called variable record equalization, has disappeared with modern records, which are standardized on the RIAA recording characteristic.

The control section of the amplifier allows the operator to regulate the volume, and, in most cases, to either accentuate or attenuate ("boost" or "cut") the bass and treble portions of the reproduced sound independently. The primary function of tone control is to compensate for deficiencies in associated equipment or program material, and to compensate for acoustical conditions of the room in which the music is heard. When the control section and phonograph preamplifier are combined on one chassis, the entire unit is commonly referred to as a preamplifier.

The power amplifier receives the electrical signal as it is finally shaped, and releases another signal, ideally identical in all respects except power. The power amplification may be tens of millions of times, from a fraction of a microwatt (one millionth of a watt) to dozens of watts.

Although the demands on the amplifier are very great, and although it appears to be the most complicated of the system components, it is the least imperfect of these components. The percentages of harmonic and intermodulation distortion, the irregularities of frequency response, and the extraneous noise introduced by an amplifier built according to the best current design practice, and without regard for cost, are such that they are not limiting factors in the fidelity of the reproduced sound.

The final component of the sound system is the loudspeaker system, which consists of the speaker mechanism itself and the speaker enclosure. The loudspeaker converts the alternating electrical output of the amplifier into mechanical vibrations of a cone or diaphragm. But the cone vibrating by itself cannot, for reasons that will be discussed further on, produce adequate bass energy. It must be mounted in an enclosure or baffle of some sort, which gives the vibrating surface the "bite" of air that it needs to radiate low-frequency sound.

The speaker and its enclosure, like the amplifier, should introduce as little distortion and frequency irregularity into the signal as possible. Typical speaker deficiencies are irregular frequency response, poor transient response (hangover), and harmonic and intermodulation distortion.

Two other components are shown in *Fig.* 3–4. The *tuner* is a device which converts AM or FM radio signals to audio signals that can be handled by the audio amplifier; the tape transport mechanism, with its associated pream-

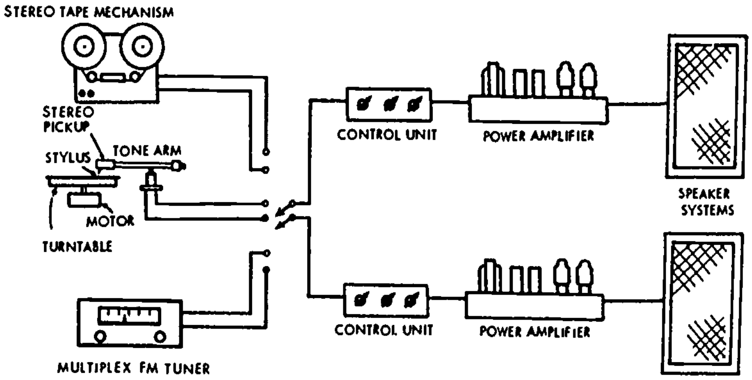


Fig. 3-5. A stereo reproducing system.

plifier, provides a signal of the same nature as that coming from the tuner or phonograph pickup.

Fig. 3-5 shows the basic elements of a stereo reproducing system. The stereo tape mechanism has two heads which independently reproduce each channel that is recorded in parallel on the tape; the stereo pickup provides two separate output signals from the two channels recorded in the groove (the turntable and pickup arm do not have to be duplicated); the stereo tuner receives the "multiplex" FM stereo signal and separates it into two separate channels, which it feeds independently to each of the control units. Each control unit and each power output is shown duplicated. The two control units and power amplifiers may be separate, or they may be combined on one chassis, or all four units may be combined on one chassis, but in any case they must provide independent amplification for each channel.

10 The Future of Direct Current Power Transmission

N. L. Allen

A popular article published in 1967.

The history of technology provides many examples of unexpected turns of fortune, and electrical technology is no exception. It frequently happens that a principle or technique, originally the basis of a well-established system, is superseded by a device making a significant advance, only to reappear in a different guise as the 'last word' in the state of the art. An obvious example is the crystal of the early radio receiver. This was superseded by the thermionic valve, but it has now developed into the more sophisticated form of the transistor. Not many years before the era of the crystal receiver, an appreciable proportion of electrical energy was generated, transmitted, and used in the form of direct current. At that time, generation and consumption usually took place in the same locality, distribution was simple, and the quantities of energy transmitted were small by modern standards. However, serious limitations appeared as it became necessary to distribute electrical energy more widely, and direct current as the distributing medium gave way to alternating current.

In many countries, the economic advantages of being able to concentrate power generation in large stations have led to the adoption of a comprehensive network of power lines that interconnect generating plant and the areas where the power is used. As the length of a power line increases, the current passed, for minimum power loss, decreases: the economic operating voltage for transmission of a given power therefore increases. The transmission of larger quantities of energy at high voltages and low currents is greatly facilitated by the ease with which alternating current can be transformed to the voltage most appropriate for the power lines. In the receiving areas of the system, the voltage can equally easily be transformed to lower values suitable for distribution, and a system of far greater flexibility can be set up than is the case with direct current. Further, it is difficult to switch and, particularly, to interrupt direct current. The interruption of an alternating current by circuit breakers is relatively easy because the current passes through zero twice in every cycle.

This combination of circumstances made alternating current the natural choice as power systems increased in size. The main links operated initially at 132 kilovolts, but the need for increased power during the post-war years has led to the adoption of 275 kilovolts and, more recently, 400 kilovolts as the operating voltages of the principal links in Britain. The power is distributed locally at lower voltages. During this period, the remaining direct current distribution systems have been reduced or eliminated.

Transmission over long distances

What, then, is the place of direct current? There is certainly no good reason for turning away completely from alternating current distribution. But there have always been some situations in power distribution practice in which direct current has distinct advantages over alternating current, and it is worth while considering what these situations are.

One basic factor in power system design is the need to find the simplest and most efficient means of transferring power from one point to another. Figure 1(a) shows the basic three-phase alternating current system and figure 1(b) a favoured direct current system, which has positive and negative polarities on the two lines, and is linked by convertors to alternating current for generation at one end and distribution at the other. In both cases, the maximum voltage to earth is E, but for alternating current, it is the root-mean-square value $E/\sqrt{2}$ that determines the power transmitted. This is $3EI_A \cos \varphi/\sqrt{2}$, where I_A is the current in each conductor, lagging behind the voltage in phase by φ degrees. In the direct current system, the power transmitted by each line is EI_D, where I_D is the current. For transmission of equal power by the two systems, therefore, it can be shown that each alternating current line has $4/(3 \cos^2 \varphi)$ times the cross sectional area of the corresponding direct current line, a factor which is always greater than 1·33. Moreover, the alternating current system requires three cables rather than two, so that the amount of copper required is $2/\cos^2 \varphi$ times that in the direct current system, a factor which is always greater than 2.

Direct current, then, reduces the cost of the cables. This may appear trivial compared with the other capital costs in electrical systems, but over great distances, as in the United States and the Soviet Union, the saving in cable, and in the means of supporting the cable, becomes a very significant factor that can outweigh the cost of providing the convertor stations at each end of the system.

Great distances bring further problems in alternating current transmission that do not occur with direct current. These problems arise from the relationship between the wavelength of the oscillation and the dimensions of the system. The quarter-wavelength of a 50 cycles per second wave in air is about 900 miles, and the transmission of energy through a conductor can be regarded as due to an influx of energy along its length from the electromagnetic field that surrounds it. Over short distances, this field is very nearly the same at all points, since electromagnetic energy is conveyed with the velocity of light. But at distances greater than 900 miles, the fact that the velocity of light is finite results in significant differences, at any instant, in the phase of the current at the two ends.

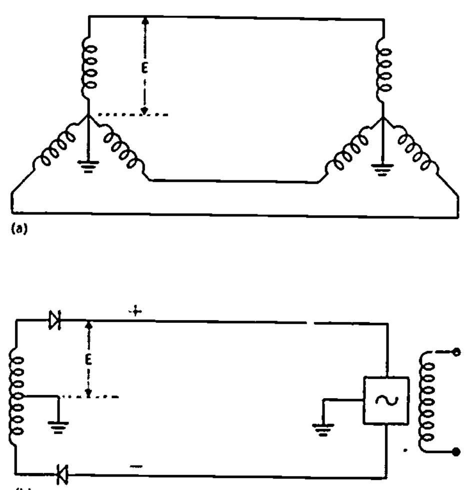


Figure 1 Simplified distribution systems: (a) alternating current, (b) direct current.

This situation leads to difficulties where two parts of a power circuit, joined by a long alternating current link, are out of phase and where a loop is formed through another part of the network of different length. Large circulating currents will be set up unless some form of compensation is applied. A direct current link obviates these difficulties; as a corollary, it may be noted also that if a direct current line is used to link two alternating current systems, they need not be synchronized with each other.

Transmission over short distances

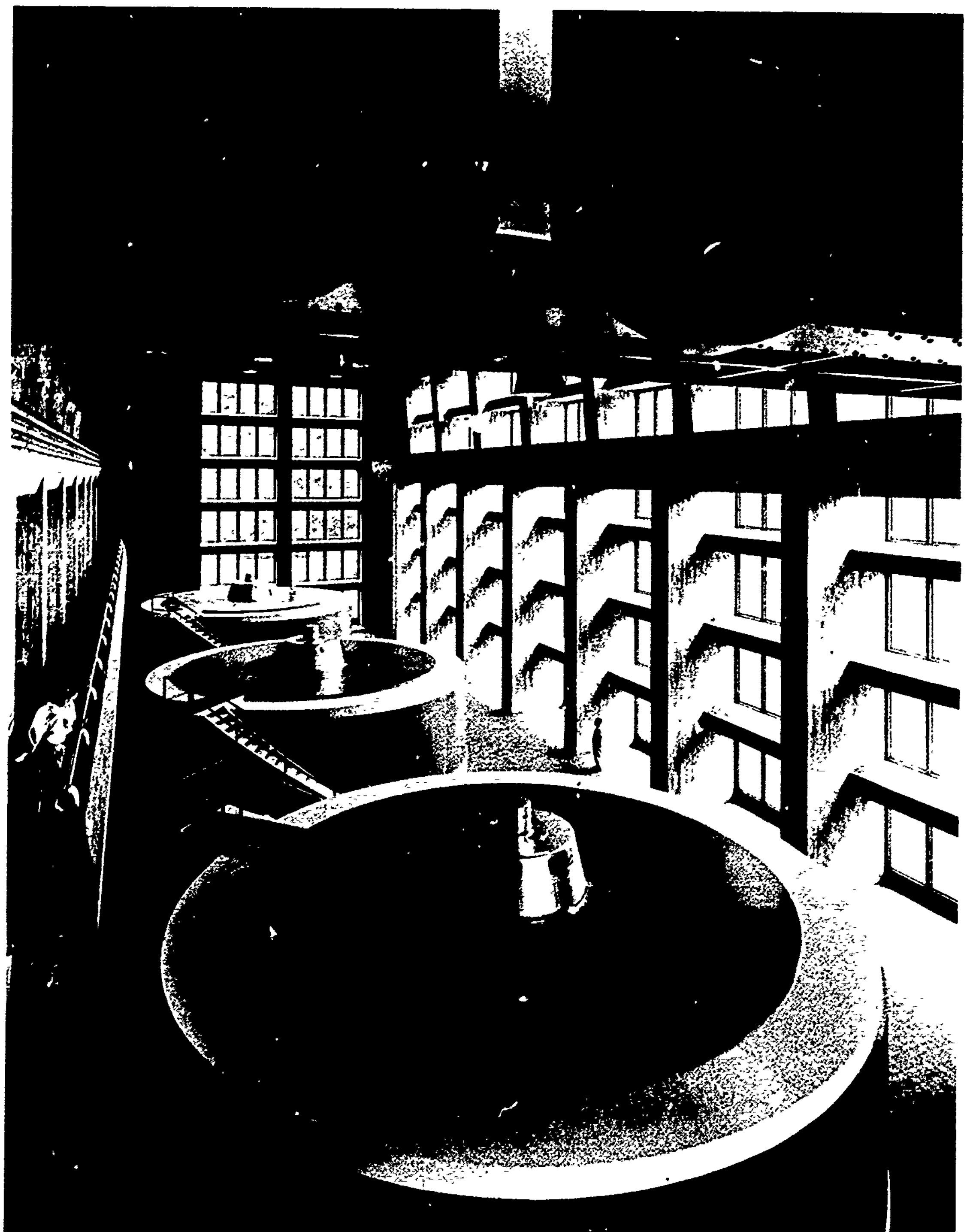
For long-distance transmission, overhead lines, supported by towers, are used. The virtues of direct current are most clearly shown when the current is carried by underground or underwater cables. Here, the central core of the cable, which is at the transmission voltage, is surrounded by an insulant, the exterior of which is at earth potential. This constitutes a coaxial capacitor, and the capacitance per mile of a cable rated at 200 kV is typically about 0·3 microfarads. In an alternating current circuit, this capacitance is charged and discharged, through the inductance and resistance of the cable itself, once every half-cycle. Additional generating capacity is needed to supply this charging current. In the example quoted, at 200 kV, the charging current requires about 5000 kilovolt-amperes per mile of cable [1, 2]; at 400 kV the figure is about 15 000 kilovolt-amperes per mile. For appreciable lengths of cable, the losses become such that the charging currents must be supplied at intermediate points. At 200 kV, these points are about 25 miles apart for 50-cycle alternating current; at 400 kV, only 15 miles. Thus, alternating current transmission becomes impracticable in cables over long distances. Further, the cost of the generating capacity needed to supply the charging current is significant. Taking a rough figure of £50 per kilowatt of installed capacity at the generating station, this extra cost is £250 000 per mile for a 200 kilovolt cable. By contrast, with direct current in the steady state, there is no charging current. It may well be worthwhile, therefore, to accept the cost of converting to direct current to avoid having to provide this charging current. Direct current is also advantageous in that there are no dielectric losses due to reversal of the electric stress in the insulant.

The balance between the two systems

To summarize, direct current has significant advantages for the transmission of bulk power over great distances by overhead lines, and over short or long distances by cable. In addition to the technical advantages already examined, direct current may be valuable in linking two alternating current systems that need not then be synchronized. Alternatively, a very large alternating current system may be divided by direct current links into two or more smaller systems: this is a possible future development as power systems continue to increase in size. It is necessary, however, to examine some disadvantages of direct current, and some relevant non-technical factors, to demonstrate the balance affecting the final choice of system.

The most obvious drawback to the use of direct current is the need for conversion at each end of the link in order to integrate it with established alternating current systems. The technical details are outlined later, but it may be mentioned here that the cost of the conversion equipment is about twice that of the alternating current equipment required for the termination of a power line of corresponding size and output [3]. These costs must be set against the savings inherent in the direct current system. There is therefore, a limit to the length of a line, below which the capital outlay on a direct current system is higher than that of an alternating current system. Estimates of the critical length for a long overhead line naturally vary, depending mainly on the power to be transmitted and the voltage to be employed, but figures of more than 300 miles have frequently been quoted [4]. This approach is unlikely to be favoured, therefore, in the British Isles, but such systems are being developed in the United States and in the Soviet Union. For underground or submarine cables, where dielectric losses and charging currents are so important, the 'critical length' is reduced to about 30 miles, and it is in short submarine links and in urban transmission lines that direct current finds its second important application. Indeed, where large amounts of power have to be introduced into large cities, legal and social considerations may predominate over technical and economic factors. It is frequently extremely difficult to obtain permission to erect overhead lines in urban areas, and the disturbance to local amenities caused by the towers for high-tension cables may not be justifiable. Underground cables become necessary, and it is preferable to use direct current for distances greater than about 30 miles.

In choosing between the systems, the fact that there can be no direct current transformer and that there is no satisfactory circuit breaker ensures that alternating current maintains its general superiority for distribution purposes. The use of direct current is thus confined to the bulk transmission of high power between discrete parts of a system or between two separate systems.



Electric generators at Fontana Dam, North Carolina.

James Clerk Maxwell, Part II

James R. Newman

A biographical essay published in 1955.

In February, 1858, Maxwell wrote a letter to his aunt, Miss Cay, beginning, "This comes to tell you that I am going to have a wife." "Don't be afraid," he added, "she is not mathematical, but there are other things besides that, and she certainly won't stop mathematics." His engagement to Katherine Mary Dewar, daughter of the principal of Marischal College, was formally announced the same month, and in June they were married.

Their union became very close: they enjoyed doing things together — horseback riding, reading aloud to each other, traveling — and he even found useful tasks for her in his experimental work. The marriage was childless, but this very fact increased the couple's dependency and devotion. Maxwell regarded the marriage tie in an "almost mystical manner." The published letters to his wife overflow with religiosity.*

The Aberdeen appointment terminated in 1860 when the two colleges, King's and Marischal, were fused into a new university and Maxwell's chair in physics at Marischal was eliminated. He was not long at liberty. In the summer of the same year he became professor of natural philosophy at King's College, London, a post he retained until 1865. The teaching schedule at King's was long and arduous; in the evenings there were lectures to be given to "artisans" as part of his regular duties. Living in London offered him the opportunity to see something of Faraday, with whom, up to this time, Maxwell had had only correspondence, to make the acquaintance of other scientific men and to renew old friendships. He was no solitary. "Work is good, and reading is good, but friends are better," he wrote to his friend Litchfield.

Yet despite academic duties and social distractions, the five years in London were the most productive of his life. The paper "On the Theory of Three Primary Colors," the two articles in the *Philosophical Magazine* on "Physical Lines of Force" and the culminating electrical memoir "A Dynamical Theory of the Electromagnetic Field," the Bakerian lecture "On the Viscosity or Internal Friction of Air and other Gases," and the celebrated paper "On the Dynamical Theory of Gases," all belong to this period. He also performed important experimental work during these years. At his house in Kensington,

* He did not write in this vein to others and it is a little puzzling why he found it necessary in corresponding with her to quote Scriptures, to express the fervent hope that the Lord would protect her from evil, and that she would get her eyes off "things seen and temporal and be refreshed with things eternal."

in a large garret, he measured the viscosity of gases and obtained practical confirmation of the theoretical work I have described. (For example, he found that the viscosity of air at 12 millimeters of mercury measured the same as at normal pressure of 760 millimeters, thus proving that viscosity is independent of density.) To maintain the necessary temperature, a fire had to be kept up in the midst of very hot weather and kettles kept boiling to produce steam, which would be allowed to flow into the room. Mrs. Maxwell acted as stoker. Another investigation dealt with the ratio of the electromagnetic to the electrostatic unit of electricity and led to one of Maxwell's greatest discoveries. But I must postpone explaining this work, even though to do so means abandoning the strict chronology of events in Maxwell's life, until I have sketched the development of his ideas on electricity.

To gain an appreciation of Maxwell's stupendous contribution to this branch of science it is necessary first to describe very briefly the position of electrical theory when he embarked on his studies.

In the eighteenth century, Charles Augustin de Coulomb established the fundamental facts of electrostatic attraction and repulsion. He showed that an inverse-square law, resembling that of gravitational forces, applied to electric charges: attraction or repulsion between charged bodies is directly proportional to the product of the charges and inversely proportional to the square of the distance between them.* (The same discoveries, and others going beyond them, were made earlier by the brilliant English recluse Henry Cavendish, but his researches remained unpublished until 1879.) The next major advance was that of Hans Oersted, who in 1819 found that the flow of electric current through a wire parallel to a magnetic needle makes the needle swing to a position at right

* $F = k\frac{qq'}{r^2}$, where F equals the force; k, a constant; q and q', the charges; r, the separating distance.

angles to the current. In other words, a current produces a magnetic field.

A complementary series of advances was made early in the same century by the French physicist and mathematician André Ampère, whom Maxwell called the Newton of electricity. The accolade was not undeserved, but there is a special reason for Maxwell's conferring it: Ampère was the first to explain the relationship of electric currents in terms of mechanical action,* an approach later perfected by Maxwell himself. By experiment Ampère learned that a coil of wire carrying an electric current behaves like a magnet, and he suggested that a magnet owes its property to tiny electrical currents inside the steel molecules. Thus a great conceptual link was forged; for magnetism was shown to be not distinct from electricity, but rather a name we give to some of the effects of moving electric currents.

The crown of these fundamental researches was the immortal work of Michael Faraday. He found that an electric current flowing in one circuit can cause ("induce") a current to flow in another circuit; that there is a magnetic field between two currents; that a current can also be induced to flow in a wire by use of a magnet — in other words, as a symmetric counterpart to the phenomena discovered by Oersted and Ampère, that changes in a magnetic field produce an electric current.

Faraday's explanation of these phenomena is of central importance to understanding Maxwell's work. He imagined *lines of force* running through space as the instrumentality of electric and magnetic actions.

These lines, it should be emphasized, were not conceived as mere mathematical makeshifts, but as entities possessing physical properties. The lines spread out in every direction from an electric charge or magnetic pole; every electric line of force

* He showed how to calculate the mechanical forces between circuits carrying currents, from an assumed law of force between each pair of elements of the circuit.

starts from a positive charge and ends on a negative charge; the more powerful the source, the more lines emanate from it. Along the lines there is tension, a kind of pull, so that each line behaves like an elastic thread trying to shorten itself; lines of force repel each other sideways; the ends of a line of force, representing charges, can move freely over the surface of a conductor but are anchored on an insulator.

This hypothetical system, for which Faraday was convinced he had found experimental evidence, was the starting point of Maxwell's studies. He believed in it; he sought to develop it.

However, it must not be supposed that everyone accepted Faraday's hypothesis. In fact, the majority of electricians — I use the term in its older sense — regarded lines of force as a concept much inferior to that of "action at a distance." They likened electricity to gravitation. They imagined a charge (or mass) situated at one point in space mysteriously influencing a charge (or mass) at another point, with no linkage or connection of any kind, however tenuous, bridging the distance between the charges (or masses). Where Faraday sought to assimilate the behavior of electricity to that of a mechanical system, in which all parts are joined by levers, pulleys, ropes and so on, the others held electricity to be a special case, to which mechanical analogies were inapplicable. Maxwell admirably summarized the cleavage between the two views: "Faraday, in his mind's eye, saw lines of force traversing all space, where the mathematicians saw centres of force attracting at a distance; Faraday saw a medium where they saw nothing but distance; Faraday sought the seat of the phenomena in real actions going on in the medium, they were satisfied that they had found it in a power of action at a distance impressed on the electric fluids."

Maxwell's first electrical paper "On Faraday's Lines of Force" was delivered at Cambridge in 1855, within a few months after he had finished reading Faraday's *Experimental Researches*. What he tried to do was imagine a physical model embodying Faraday's lines, whose behavior, like that of any machine, could be reduced to formulae and numbers. He did not suggest that the model represented the actual state of things; on the other hand, he had no confidence in what mathematical manipulations alone would reveal about the actual state of things. It was important, he said, so to balance the method of investigation that the mind at every step is permitted "to lay hold of a clear physical conception, without being committed to any theory founded on the physical science from which that conception is borrowed."* Such a method will neither lead

* The opening paragraph of the paper is worth giving in full. "The present state of electrical science seems peculiarly unfavorable to speculation. The laws of the distribution of electricity on the surface of conductors have been analytically deduced from experiment; some parts of the mathematical theory of magnetism are established, while in other parts the experimental data are wanting; the theory of the conduction of galvanism and that of the mutual attraction of conductors have been reduced to mathematical formulae, but have not fallen into relation with the other parts of the science. No electrical theory can now be put forth, unless it shows the connection not only between electricity at rest and current electricity, but between the attractions and inductive effects of electricity in both states. Such a theory must accurately satisfy those laws the mathematical form of which is known, and must afford the means of calculating the effects in the limiting cases where the known formulae are inapplicable. In order therefore to appreciate the requirements of the science, the student must make himself familiar with a considerable body of most intricate mathematics, the mere attention of which in the memory materially interferes with further progress. The first process therefore in the effectual study of the science, must be one of simplification and reduction of the results of previous investigation to a form in which the mind can grasp them. The results of this simplification may take the form of a purely mathematical formula or of a physical hypothesis. In the first case we entirely lose sight of the phenomena to be explained; and though we may trace out the consequences of given laws, we can never obtain more extended views of the connections of the subject. If, on the other hand, we adopt a physical hypothesis, we see the phenomena only through a medium, and are liable to that blindness to facts and rashness in assumption which a partial explanation encourages. We must therefore discover some method of investigation which allows the mind at every step to lay hold of a clear physical conception, without being committed to any theory founded on the physical science from which that conception is borrowed, so that it is neither drawn aside from the subject in pursuit of analytical subtleties, nor carried beyond the truth by a favorite hypothesis. In order to obtain physical ideas without adopting a physical theory we must make ourselves familiar with the existence of physical analogies. By a physical analogy I mean that partial similarity between the laws of one science and those of another which makes each of them illustrate the other. Thus all the mathematical sciences are founded on relations between physical laws and laws of numbers, so that the aim of exact science is to reduce the problems of nature to the determination of quantities by operations with numbers."

into a blind alley of abstractions, nor permit the investigator to be "carried beyond the truth by a favorite hypothesis."

Analogies are, of course, the lifeblood of scientific speculation. Maxwell gives a number of examples, among them the illuminating suggestion of William Thomson comparing the formulae of the motion of heat with those of attractions (such as gravitation and electricity) varying inversely as the square of the distance. To be sure, the quantities entering into heat formulae — temperature, flow of heat, conductivity — are distinct from a quantity such as force entering into the formulae of inverse-square attraction. Yet the mathematical laws of both classes of phenomena are identical in form. "We have only to substitute *source of heat* for *center of attraction*, *flow of heat* for *accelerating effect of attraction* at any point, and *temperature* for *potential*, and the solution of a problem in attractions is transformed into that of a problem of heat."*

Maxwell proposed a hydrodynamical analogy to bring before the mind in "convenient and manageable form those mathematical ideas which are necessary to the study of the phenomenon of electricity."† The analogy was combined with Faraday's lines of force, so that they were converted from lines into "tubes of flow" carrying an incompressible fluid such as water. He was then able to show that the steady flow of particles of this fluid would give rise to tensions and pressures corresponding to electrical effects. The fluid moving through a system of such tubes represented electricity in motion; the form and diameter of the tubes gave information as to strength and direction of fluid (electric) flow. The velocity of the fluid was the equivalent of electrical force; differences of fluid pressure were analogous to differences of electrical pressure or potential. Since the tubes were flexible and elastic, and ar-

* "On Faraday's Lines of Force," *Transactions of the Cambridge Philosophical Society*, vol. X, part I, included in *The Scientific Papers of James Clerk Maxwell, op. cit.*

† *Ibid.*

ranged so as to form surfaces — each tube being in contact with its neighbors — pressure transmitted from tube to tube furnished an analogy to electrical induction.

One of Faraday's key concepts deals with the effect on space of lines of magnetic force. A wire introduced into ordinary space remains inert; but if magnetic lines of force are introduced into the space, a current flows through the wire. Faraday explained this by saying that the introduction of the magnet threw the space into an "electro-tonic state." This concept could not be fitted into the hydrodynamical analogy; Maxwell admitted that while he could handle Faraday's conjecture mathematically, the electro-tonic state at any point of space being defined "as a quantity determinate in magnitude and direction," his representation involved no physical theory — "it is only a kind of artificial notation."*

It was a wonderful paper, and Faraday, to whom Maxwell sent a copy, appreciated how much it advanced the "interests of philosophical truth." "I was at first almost frightened," he wrote Maxwell, "when I saw such mathematical force made to bear upon the subject, and then wondered to see that the subject stood it so well."† Other students, however, thought the subject stood it not at all well. Electricity was mysterious enough without adding tubes and surfaces and incompressible fluids. But Maxwell, who had good training in being considered queer, went on with the task of extending Faraday's ideas.

The second great memoir, *On Physical Lines of Force*, a series of three papers published in the *Philosophical Magazine* in 1861 and 1862, was an attempt to describe a more elaborate mechanism that would not only account for electrostatic effects but also explain magnetic attraction and Faraday's concept of

* For a discussion of Maxwell's use of physical analogy, see Joseph Turner, "Maxwell on the Method of Physical Analogy," *The British Journal for the Philosophy of Science* vol. VI, no. 23, November, 1955.

† Campbell and Garnett, *op. cit.*, p. 519.

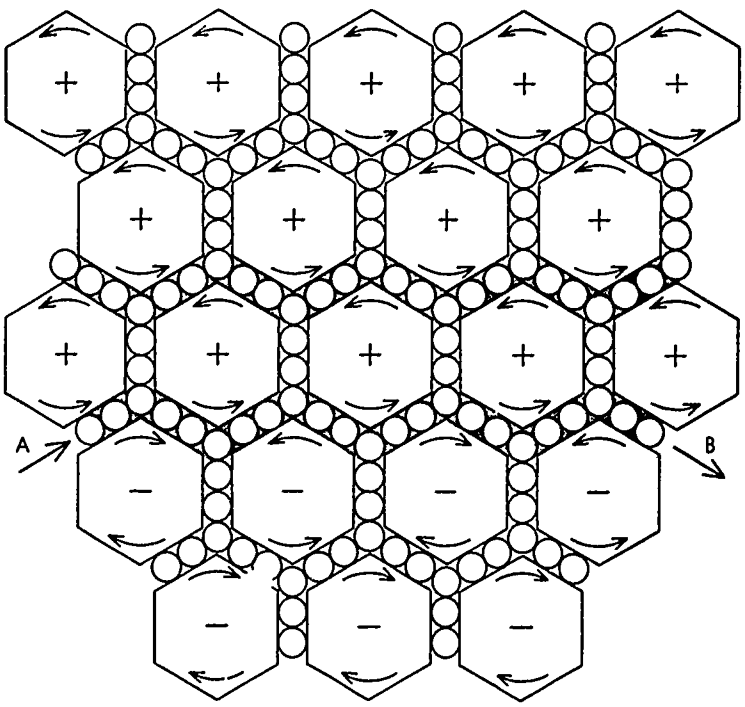
electromagnetic induction. Again, Maxwell used a concrete, mechanical image to exhibit and develop his theory.* For, as he said, "scientific truth should be regarded as equally scientific whether it appears in the robust form and vivid colouring of a physical illu ion or in the tenuity and paleness of a symbolic expression.

In the new model a magnetic field is produced by the rotation in space of what Maxwell called "molecular vortices." These may be thought of as slender cylinders (Maxwell himself had a disconcerting way of modifying the image as he went along) that rotate round the lines of magnetic force. The lines, traced by the pattern of iron filings about a magnet, are the axes of rotation of the cylinders; the velocity of rotation depends on the intensity of the magnetic force. Two mechanical effects are associated with the cylinders: *tension* in the direction of the lines of force, and *pressure*, exerted in the "equatorial" direction (i.e., lateral pressure), arising from the centrifugal force produced by the rotating cylinders. Combined, these effects mechanically reproduce magnetic phenomena: magnetism is a force exerted both along the axis and outward from the axis.

It may now be asked how this curious arrangement fitted in with the known facts that an electric current produces a magnetic field, and changing magnetic forces produce an electric current. Step by step Maxwell answers this question.

The first point to clarify is the structure of a uniform magnetic field. Maxwell supposed this to consist of a portion of space filled with cylinders rotating at the same velocity and in the same direction "about axes nearly parallel." But immediately a puzzle confronted him. Since the cylinders are in contact, how can they possibly rotate in the same direction? For

* As Turner (*op. cit.*) points out Maxwell employed two analogies. One bridged a stationary field and a solid under stress. The other is between electricity and fluid motion, "with its suggestion that Ampère's laws be modified to satisfy the equation of continuity."



Model of an electromagnetic field used by Maxwell visualized "Molecular vortices" rotating in space. In this illustration the vortices are slender cylinders seen from the end. (Maxwell gave the cylinders a hexagonal cross section to simplify the geometry of the model.) Between the vortices are small "idle wheels." If a row of the idle wheels is moved from A *toward* B, *they cause the adjacent vortices to rotate in the opposite direction.* (Scientific American)

as everyone knows, a revolving wheel or cylinder causes its neighbor to revolve in the opposite direction; thus one would expect the rotation of the cylinders to alternate in direction from one to the next. Maxwell hit upon a pretty idea. He supposed the cylinders to be separated by rows of small spheres, like layers of ball bearings, which acted as gears (in Maxwell's words, "idle wheels"). This arrangement, resembling a device envisaged a century earlier by John Bernoulli, the younger, fulfilled the requirement. The spheres rotate in an opposite sense to that of each of the cylinders with which they are in contact, and so the cylinders all rotate in the same direction.

And now, as just reward for his ingenuity, Maxwell found that the spheres could be made to serve another, even more valuable, purpose. Think of them as particles of electricity. Then by purely mechanical reasoning it can be shown that their motions in the machine of which they are a part serve to explain many electrical phenomena.

Consider these examples. In an unchanging magnetic field the cylinders all rotate at the same constant rate; thus they maintain a constant magnetic field. The little rotating spheres keep their position; there is no flow of particles, hence no electric current, a result that tallies with the properties of a uniform magnetic field. Now suppose a change in the magnetic force. This means a change in the velocity of rotation of the cylinders. As each cylinder is speeded up, it transmits the change in velocity to its neighbors. But since a cylinder now rotates at a slightly different speed from its neighbor, the spheres between them are torn from their positions by a kind of shearing action. In other words, they begin to move from their centers of rotation, in addition to rotating. This motion of translation is an electric current; again, a result that tallies with the properties of a changing magnetic field.

Observe now how the model begins to live a life of its own. It was designed, as J. J. Thomson has pointed out,* to exhibit Faraday's great discovery that magnetic changes produce electric currents. It suggested to Maxwell the no less striking converse phenomenon that changes in electric force might produce magnetism.† Assume the spheres and cylinders are at rest. If

* Sir J. J. Thompson. "James Clerk Maxwell," in *James Clerk Maxwell, A Commemoration Volume, op. cit.*

† Ampère, of course, had already demonstrated that currents *in wires* produced accompanying magnetic fields.

a force is applied to the spheres of electricity, they begin to rotate, causing the cylinders of magnetism with which they are in contact to rotate in the opposite direction. The rotation of the cylinders indicates a magnetic force. Moreover, the model holds up even as to details. Take a single illustration. Magnetism acts at right angles to the direction of flow of current. If you will examine the diagram of Maxwell's model, you will see that the cylinders will rotate in the direction perpendicular to the motion of the spheres, thus bearing out the observation that a magnetic force acts at right angles to the flow of a current.

"I do not bring it forward," Maxwell wrote of his system, "as a mode of connection existing in Nature. . . . It is, however, a mode of connection which is mechanically conceivable and easily investigated, and it serves to bring out the actual mechanical connection between the known electromagnetic phenomena.* Certain aspects of these "mechanical connections" have already been discussed — rotations, pressures, tensions — which account for the reciprocal relations between currents and magnetic forces.† The connections also serve to explain the repulsion between two parallel wires carrying currents in opposite directions, an effect produced by the centrifugal pressures of the revolving cylinders acting on the electrical particles between them. The induction of currents is similarly elucidated: this phenomenon, says Maxwell, is simply "part of the process of communicating the rotary velocity of the vortices [cylinders] from one part of the field to another." In other words, whenever electricity (Maxwell's particles) "yields to an electromotive force," induced currents result. His diagram and the accompanying text make this beautifully clear.

Maxwell was not done with his model. It had helped in the

* "On Physical Lines of Force," *op. cit.*

† The model explained, for example, why a current of electricity generated heat: for as the particles (or spheres) move from one cylinder to another, "they experience resistance, and generate irregular motions, which constitute heat."

interpretation of magnetism, the behavior of electric currents, the phenomenon of induction; it had yet to pass the supreme test: that is, to supply a mechanical explanation of the origin of electromagnetic waves. To orient ourselves in this matter we must examine briefly the question of condensers and insulators.

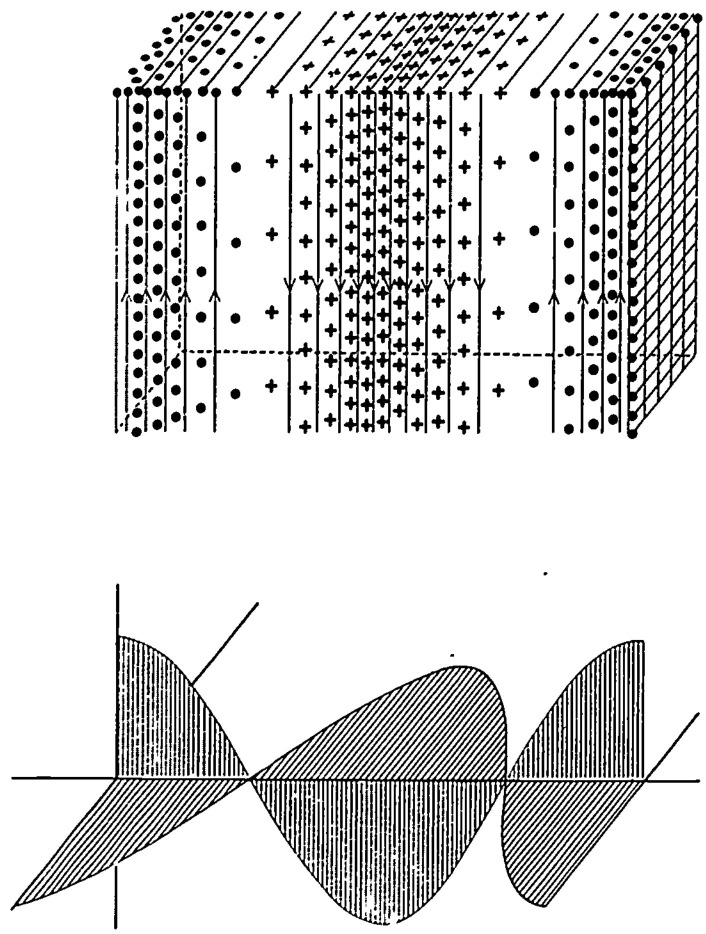
An electric condenser is a device for storing electricity. In its simplest form it consists of two conducting plates separated by an insulating material, or dielectric as it is called. The plates can be charged, after which the charges attract each other through the dielectric and are thus said to be "bound" in place. Faraday in his experiments had come upon a curious fact. He found that two condensers of the same size, fed by the same electric source and with insulation of equal thickness, differed markedly in their capacity to take or to hold a charge if the insulating material (dielectric) was different. This was difficult to understand if all dielectrics were equally impermeable to an electric current. Moreover, if it were true, as Maxwell already was beginning to suspect, that light itself is an electrical phenomenon, how could light pass through certain dielectrics, empty space among them? With the help of his model, Maxwell advanced a bold hypothesis. Conductors, he said, pass a current when the electrical particles they contain are acted upon by an electric force. Under such an impulsion, the little particles move more or less freely from cylinder to cylinder, and the current flows as long as the force persists. Not so in a dielectric. The particles are present but an easy passage from cylinder to cylinder is impossible. This fact may be taken as the characteristic attribute of a dielectric, having to do with its physical structure. Yet it was known that "localized electric phenomena do occur in dielectrics." Maxwell suggested that these phenomena also are currents — but of a special kind. When an electric force acts on a dielectric, the particles of electricity are "displaced," but not entirely torn loose; that is, they behave like a ship riding at anchor in a storm. The medium in which they are located, the magnetic cylinders, is elastic; under pressure the particles move, a limited distance, until the force pushing them is balanced by the stresses due to the elastic reaction of the medium. Thus a state of equilibrium is attained. But as soon as the impelling force ceases to act, the particles snap back to their original positions. The net effect of these mechanical actions is twofold. First, the initial displacement of the electric particles constitutes a current that passes through the dielectric. A current of this type is called a *displacement current* to distinguish it from currents that flow through conductors and are therefore known as conduction currents.* Wherever there is an electric force, said Maxwell, there is displacement; wherever there is displacement, there is a current.

Second, whenever the pressure displacing the particles is released, and they snap back, they overshoot and oscillate briefly about their fixed positions. The oscillation is transmitted through the magnetic medium (the insulator) as a wave. This wave is the return phase of the displacement current.† (Maxwell suggested this disturbance on analogy to the displacement of an elastic solid under stress.)

Maxwell next arrived at an epoch-making conclusion. The velocity of the displacement wave, or current, depends on the electrical properties of the medium in which it moves. Moreover, this velocity, as he showed, was "within the limits of experimental error, the same as that of light." Hence, he in-

* The contrast between displacement currents and currents through conductors was vividly expressed by Henri Poincaré. A displacement current, he said, is an elastic reaction like the compression of a spring: it can only be effected by pressure against resistance. Equilibrium is reached when resistance balances pressure. When the pressure is removed the spring regains its original form. A conduction current, on the other hand, is like a viscous reaction such as is encountered in moving a body immersed in water. It can be effected only by pressure; the resistance depends on velocity; the motion continues as long as the motive force acts, and equilibrium will never be established. "The body does not return to the starting point, and the energy expended in moving it cannot be restored, having been completely transformed into heat through the viscosity of the water." (*Maxwell's Theory and Wireless Telegraphy*, New York, 1904.)

† If the electric force applied to the insulator is varied continually, it will produce a continually varying displacement wave: in other words, a continuing current.



Electromagnetic wave as visualized by Maxwell is a moving disturbance which tends to separate positive (plus sign) and negative (dot) charges. In the drawing at the top, magnetic lines of force (arrows) lie at right angles to the direction in which the disturbance is moving. The drawing at the bottom depicts the two components of the electromagnetic wave. The electrical component is shown in black, the magnetic component in color. (Scientific American)

ferred, "the elasticity of the magnetic medium in air is the same as that of the luminiferous medium, if these two coexistent, coextensive and equally elastic media are not rather one medium."

More must be said as to how Maxwell actually arrived at this conclusion. In the 1850s the German physicists Wilhelm Weber and Friedrich Kohlrausch had investigated an important relationship, namely, the ratios of electrostatic to electrodynamic action. The electrostatic unit of charge was defined as the repulsion between two (like) unit charges at unit distance apart. The electrodynamic unit was defined as the repulsion between two definite lengths of wire carrying currents "which may be specified by the amount of charge which travels past any point in unit time." In order to compare the repulsion between static charges with that between moving charges, a factor of proportionality must be introduced, since the units are different for static and dynamic phenomena. That is, one must determine how many positive units of electricity flowing in one wire, and negative units flowing in the other, are required to produce between the wires a repulsion quantitatively equal to that between two static units. The factor turns out to be a velocity; for since the length of the wires is fixed, and the number of units of electricity passing a given point in a given time can be measured, what the investigator must consider is the dimension length divided by time or velocity. Weber and Kohlrausch had found that the velocity of propagation of an electric disturbance along a perfectly conducting wire is close to 3×10^{10} centimeters per second. This was an astonishing coincidence, for the figure was about the same as the velocity of light as it had been determined a few years earlier by the French physicist Hippolyte Fizeau.

Kirchhoff remarked the coincidence, but did not pursue it; Maxwell did. In 1860 he attacked the problem experimentally, using an ingenious torsion balance to compare the repulsion between two static charges and two wires carrying currents. The Weber-Kohlrausch results were roughly confirmed. Also, at about the same time (he said, in fact, that the pencil and paper work was done before seeing Weber's memoir), he calculated the velocity of displacement currents in empty space or in any other dielectric. The resulting values tallied closely.

In other words, currents in a wire, displacement currents in a dielectric, and light in empty space (which of course is a dielectric) all traveled with the same velocity. With this evidence at hand, which he communicated in a letter to Faraday in 1861, Maxwell did not hesitate to assert the identity of the two phenomena — electrical disturbances and light. "We can scarcely avoid the inference," he said, "that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena."

"On Physical Lines of Force," despite its cogwheels and other gross mechanical adjuncts, may be regarded as the most brilliant of Maxwell's electrical papers. If it did not claim to give a picture of the true state of things, it was at least enormously enlightening as to how electricity and magnetism could interact in a purely mechanical relationship. Maxwell himself summarized the achievements of the theory as follows. It explained magnetic forces as the effect of the centrifugal force of the cylinders; induction as the effect of the forces called into play when there is a change of angular velocity of the cylinders; electromotive force as an effect produced by stress on the connecting mechanism; electric displacement as a result of the elastic yielding of the mechanism; electromagnetic waves as an accompaniment of displacement. The paper is one of the rare examples of scientific literature in which one may glimpse the play of imagination, the actual exercise of inductive power, the cultivation of nascent ideas.

None of the basic concepts unfolded in this memoir was discarded in the more mathematical writings that followed. But Maxwell now had to outgrow his model. In "A Dynamical Theory of the Electromagnetic Field," published in 1864,* Maxwell, in Sir Edmund Whittaker's words, displayed the architecture of his system "stripped of the scaffolding by aid of which it had first been erected."† The particles and cylinders

* *Royal Society Transactions*, vol. CLV.

† *History of the Theories of Aether and Electricity: The Classical Theories*, London, 1951.

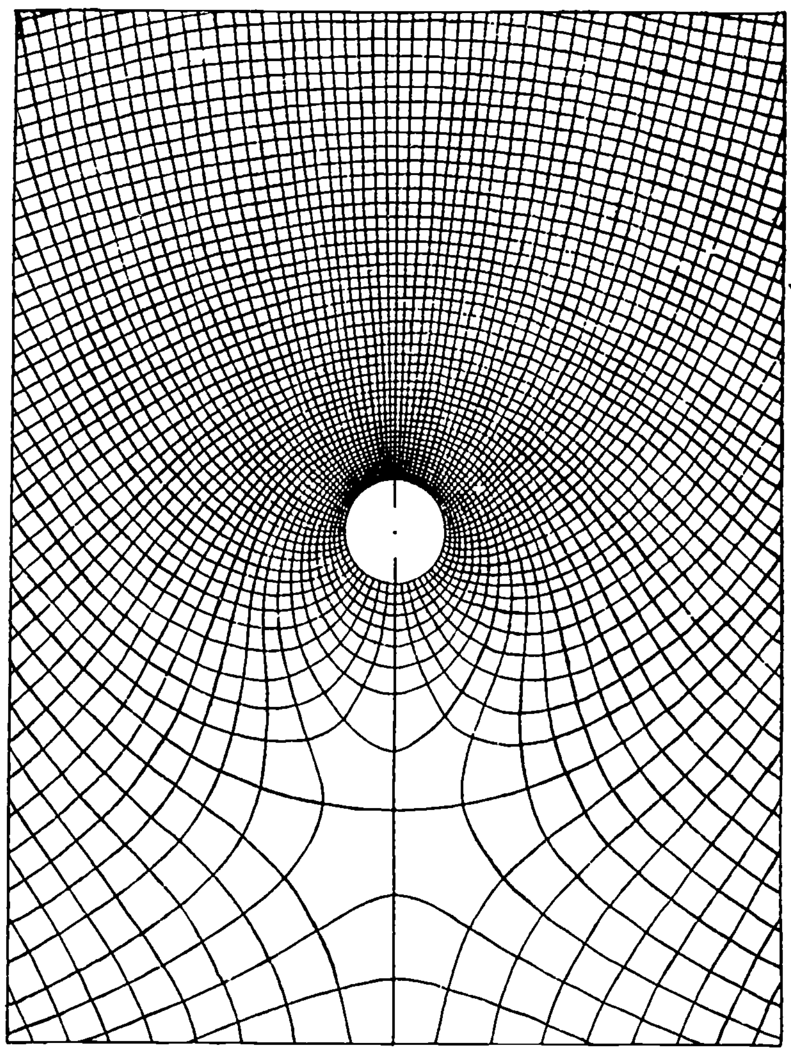
are gone; in their place is the field — "the space in the neighborhood of the electric or magnetic bodies" — and the aether, a special kind of "matter in motion by which the observed electromagnetic phenomena are produced." The matter composing the aether has marvelous properties. It is very fine and capable of permeating bodies; it fills space, is elastic and is the vehicle of "the undulations of light and heat." Yet for all its refinements and subtleties, the medium is no less a mechanical rig than the cylinders and spheres of its predecessor. It can move, transmit motions, undergo elastic deformations, store potential (mechanical) energy and release it when the deforming pressures are removed. Though susceptible to the action of electric currents and magnets, it is nonetheless a mechanism that, as Maxwell said, "must be subject to the general laws of Dynamics, and we ought to be able to work out all the consequences of its motion, provided we know the form of the relation between the motions of the parts." In the preceding paper Maxwell already had devised a set of equations that described the possible mechanical basis of electrical and magnetic phenomena, and, in particular, how certain changes in electric and magnetic forces could produce electrical waves. He now elaborated the hypothesis of displacement currents and obtained the expressions that are in substance the famous Maxwellian equations of the electromagnetic field.

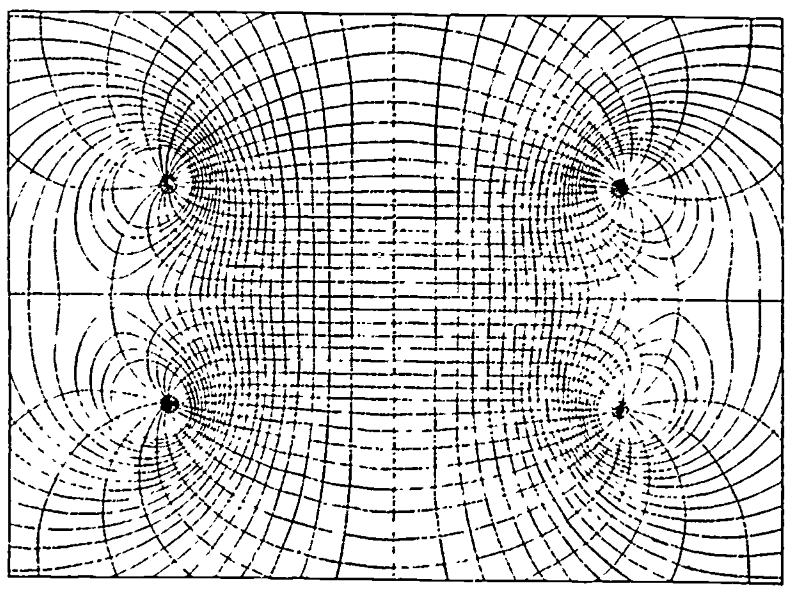
In their most finished form the equations appear in the *Treatise on Electricity and Magnetism* (1873), the culmination of Maxwell's researches, which he wrote at Glenlair in the years following his resignation from King's College. This celebrated work deals with every branch of electric and magnetic science and presents the results of twenty years of thought and experiment. Maxwell remained faithful to Faraday, whose point of view is emphasized throughout the *Treatise*. Characterizing his own part as that of an "advocate," Maxwell makes no attempt to describe the hypotheses propounded by Weber, Gauss, Riemann, Carl and Franz Neumann, or Ludwig Lorenz, the foremost cultivators of the theory of action at a distance.

The task Maxwell set himself was, first, to formulate mathematically electromagnetic phenomena as observed experimentally, and, second, to show that these mathematical relationships could be deduced from the fundamental science of dynamics; or to put it another way, that the laws of electricity in motion could be derived from the laws applicable to any system of moving bodies. As always, Maxwell was very cautious in expressing himself about the nature of electricity. It behaves, he said, like an incompressible fluid; "wherever there is electric force there is electric displacement." These, as J. J. Thomson observed, are the only definite statements about electricity to be found in the treatise, which led Hertz to say that Maxwell's theory is Maxwell's equations, and caused Helmholtz to comment that "he would be puzzled to explain what an electric charge was on Maxwell's theory beyond being the recipient of a symbol."

What are the Maxwellian equations? I cannot hope to give an easy answer to this question, but at the cost of deliberate oversimplification I must try summarily to explain them, for they are the heart of the theory.

Maxwell based the equations on four principles: (1) that an electric force acting on a conductor produces a current proportional to the force; (2) that an electric force acting on a dielectric produces displacement proportional to the force; (3) that a current produces a magnetic force (i.e., a moving electric charge is surrounded by a magnetic field) at right angles to the current's lines of flow and proportional to its intensity; (4) that a changing magnetic force (or field) produces a current proportional to the intensity of the force. The third and fourth principles exhibit a striking symmetry. The third is Faraday's law of electromagnetic induction, according to which "the rate of alteration in the number of lines of magnetic induction passing through a circuit is equal to the work done in taking unit electric charge round the circuit." Max-





Lines of force appear in Electricity and Magnetism. LEFT: *"Uniform magnetic field disturbed by an electric current in a straight conductor."* ABOVE: *"Two circular currents."* (Scientific American)

well's complementary law, the fourth principle, is that "the rate of alteration in the number of lines of electric force passing through a circuit is equal to the work done in taking a unit magnetic pole round it."

On this foundation two sets of symmetrical equations can be erected. One set expresses the continuous nature of electric and magnetic fields; the second set tells how changes in one field produce changes in the other. In these formulations the mechanical aspects of the theory are retained, perfect continuity is preserved by treating electricity as if it were an incompressible fluid, and wave phenomena are deduced as the consequences of displacement in a dielectric.

How does the concept of the field enter the theory? We have followed Maxwell as he stripped his model of its particles and cylinders and reduced it to an aetherial medium. In the *Treatise*, while not abandoning the medium altogether, he robs it of almost all its attributes other than form. The matter of the medium, as Poincare says, is left only with purely geometric properties, the atoms dwindle to mathematical points, subject to the laws of dynamics alone. The grin is left but the cat is gone. It is a perfect example of mathematical abstraction.*

The aether is a thing that wiggles when it is prodded, but does nothing on its own. An electromagnetic field consists of two kinds of energy, electrostatic or potential energy, and electrodynamic or kinetic energy. The aether, like

* Einstein made an interesting comment about Maxwell's equation and his use of the concept of the field. "He showed that the whole of what was then known about light and electromagnetic phenomena was expressed in his well-known double system of differential equations, in which the electric and the magnetic fields appear as the dependent variables. Maxwell did, indeed, try to explain, or justify, these equations by intellectual constructions. But he made use of several such constructions at the same time and took none of them really seriously, so that the equations alone appeared as the essential thing and the strength of the fields as the ultimate entities, not to be reduced to anything else. By the turn of the century the conception of the electromagnetic field as an ultimate entity had been generally accepted and serious thinkers had abandoned the belief in the justification, or the possibility, of a mechanical explanation of Clerk Maxwell's equations. Before long they were, on the contrary, actually trying to explain material points and their inertia on field theory lines with the help of Maxwell's theory, an attempt which did not, however, meet with complete success. Neglecting the important *individual* results which Clerk Maxwell's life work produced in important departments of physics, and concentrating on the changes wrought by him in our conception of the nature of physical reality, we may say this: — before Clerk Maxwell people conceived of physical reality — insofar as it is supposed to represent events in nature — as material points, whose changes consist exclusively of motions, which are subject to partial differential equations. After Maxwell they conceived physical reality as represented by continuous fields, not mechanically explicable, which are subject to partial differential equations. This change in the conception of reality is the most profound and fruitful one that has come to physics since Newton; but it has at the same time to be admitted that the program has by no means been completely carried out yet."

I am puzzled as to what Einstein meant in saying that Maxwell's equation eliminated the notion of mechanism in explaining electromagnetic phenomena. Similar views have been expressed by many other physicists and philosophers. Maxwell himself would not have agreed with this position. His writings refute it. The inference was drawn by his successors. But there is a more important a universal condenser, may be conceived as storing energy — in which case, being elastic, it is deformed. Since the aether fills all space and therefore penetrates conductors as well as dielectrics, it no longer makes any difference whether we deal with a conduction current or a displacement current; in either case the aether is set in motion. This motion is communicated mechanically from one part of the medium to the next and is apprehended by us as heat, or light, or mechanical force (as in the repulsion between wires) or other phenomena of magnetism and electricity. The ruling principle of all such phenomena, it should be observed, is that of least action. This is the grand overriding law of the parsimony of nature: every

---

point that requires clarification; namely, do the equations justify the inference? It is true that a field is not the same as a material particle, and that the motion of a particle is not the same as a change in a field. It is true also that the concept "material particle" was long held to be intuitively clear, while the concept "field" has never been so regarded. This makes it easier to say mysterious things about fields, which no one would dream of saying about particles. But a more careful definition of these concepts, as physicists actually use them, raises serious questions as to whether a field is any less suited to a "mechanistic" explanation than a system of material particles; indeed, whether a mechanistic explanation fits either or neither case. In modern physics material particles are not what they once were. They are pale abstractions, quite incapable of anything so robust as a collision. But then what is a collision? One thinks of billiard balls knocking together, as a pristine example. This, however, is a plain man's way of thinking. The modern physicist has rid his mind of such seductive images. (As far back as the eighteenth century, the Italian physicist Boscovich proposed the idea that the heart of an atom is not solid substance but a mere center of immaterial force.) As particles fade, the field becomes more substantial. Properties are now ascribed to it that make it seem more real and more potent than a billiard ball or a boulder. Of course the field is hard to describe in homely terms. Yet it is quite capable, as physicists tell us, of doing homely things. It produces and undergoes changes — now as if it were a cloud, now an engine, now an ocean. In short it has mechanical effects. By this I mean effects of a kind produced by what used to be called material particles. Moreover, it has mechanical properties. By this I mean properties of a kind produced by what we call a machine. The field can do things no system of particles or machine yet conceived can do. Since it can also do all they can do, it is a super-machine. Is there any point in saving the name? I think there is, to keep our thinking straight. We ought to keep it to describe both fields and particles or we ought to discard it entirely. If the word "mechanism" has any meaning in the universe of refined observation, it has as much meaning in relation to fields as to particles. At the same time I am quite prepared to believe that it has as little meaning in one case as the other; for that matter, no meaning in either.

action within a system is executed with the least possible expenditure of energy. It was of the first importance to Maxwell that electrical phenomena should satisfy the principle, for otherwise his mechanical explanation of the phenomena would not have been possible.

With these points in mind, we may examine a set of Maxwell's equations in a form that describes the behavior of an electromagnetic field under the most general conditions, i.e., a field moving in empty space. No conductors are present, no free charges, and the medium is a vacuum. The equations then read

$$1)\ div\,E = 0$$
$$2)\ div\,H = 0$$
$$3)\ curl\,E = -\frac{1}{c}\frac{\partial H}{\partial t}$$
$$4)\ curl\,H = \frac{1}{c}\frac{\partial E}{\partial t}$$

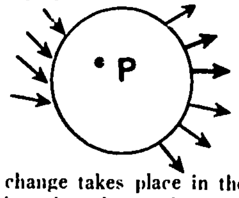
The meaning of the symbols is as follows: E and H represent electric and magnetic field strength; since they vary in time, and from place to place, they are functions of the space coordinates x, y, z (not shown) and of the time coordinate, t. C is the velocity of light and enters the equations as the rate of propagation; *div* (an abbreviation for divergence) and *curl* (an abbreviation for rotation) represent mathematical operations whose physical meaning is explained below.

Divergence is essentially a measure of rate of change. In words, then, equation 1

$$div\,E = 0$$

says that in a moving field the electric intensity is the same at every point, i.e., the rate of change is zero at every point. More loosely, this equation extends to the field the classical principle that electric lines of force can be neither created nor destroyed. Thus the equation says that the number of electric lines of force, representing the field strength, that enter any tiny volume of space must equal the number leaving it. Making use of still another analogy, if one conceives of electricity in Maxwell's idiom, as an incompressible fluid, equation 1 states that as much fluid flows out of a tiny volume of space in a given time as flows in.*

* For the reader interested in a little more detail, the following explanation may be helpful. Equation 1 states that the divergence of the electric field intensity is zero at any point in space and at any instant of time. The meaning of the equation may be visualized as follows. It is customary to represent E at a given instant of time by a series of lines whose relative density in space is proportional to E. These lines have direction because E is a vector. Consider a point P and a sphere surrounding P. Let us suppose that the intensity of the electric field on the left hemispherical surface of the sphere is uniform over the surface and is directed at each point perpendicular to the surface.



Suppose further that some change takes place in the electric field intensity E in the region occupied by the sphere but such that on the right hemispherical surface the field E is again uniform and perpendicular to the surface but stronger than on the left portion. We would indicate this increase in the intensity of E by having more lines leave the sphere on the right than enter on the left. Using the number of lines as a measure of E, we count the lines entering the spherical surface and multiply this number by the area of the hemisphere, and regard this product as negative. Let us next form the analogous product of the area and the number of lines leaving the surface, and regard this product as positive. The algebraic sum of these two products, that is, the positive plus the negative, is called the net electric flux through the spherical surface. This net flux is the divergence of E over the volume of the sphere. In our illustration the net flux of E has increased as E passes through the sphere. Hence we should say in this case that the divergence of E through the sphere is positive. If we now divide this net flux through the sphere by the volume of the sphere, we obtain the next net flux per unit volume. We now imagine that the sphere becomes smaller and smaller and contracts to the point P. Of course the net flux per unit volume changes and approaches some limiting value. This limiting value, which is a mathematical abstraction, is *div* E at the point P. Thus *div* E is essentially a measure of the spatial rate of change of E at the point P. Since equation 1 says that for electric fields *div* $E = 0$ at each point P, we may say that the net spatial rate of change of E is zero in empty space. More loosely stated, this equation says that electric field lines are neither created nor destroyed at the point P. It is to be noted that the phrase "spatial rate of change" is intended to emphasize that the divergence is concerned with the way in which E changes from point to point in space at the same instant of time. This spatial rate must be distinguished from the rate at which some quantity, for example, E itself in equation 4, may change during some interval of time.

Equation 2

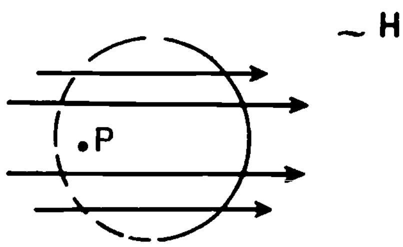
$$div\ H = 0$$

makes the same assertion for magnetic lines as equation 1 makes for electric lines.

Equation 3

$$curl\ E = -\frac{1}{c}\frac{\partial H}{\partial t}$$

is Maxwell's way of stating Faraday's law of induction. The equation describes what happens in a changing magnetic field. The right side expresses rate of change, $\frac{\partial H}{\partial t}$, multiplied by a very small factor, $-\frac{1}{c}$ (the negative sign before the fraction is purely a matter of algebraic convenience); the left side expresses the fact that an electric field is created by a changing magnetic field. But the equation is more than analytic; thanks to the sign *curl*, it actually gives a picture of the event. A simple diagram may help make this clear. Suppose the existence of a magnetic field uniform over a region of space. We draw a circle



surrounding a bundle of parallel lines, which represent the intensity and direction of the magnetic field. The circle lies in a plane perpendicular to the lines. If the field is changed (by motion or by increase or reduction of strength), it produces an electric field that acts in a circle around the lines of magnetic force (though it may also act in other directions). By summing the work done in moving unit electric charge around the circle, we obtain what is called the net electromotive force around the circle.* If the circle were made of wire, the changing magnetic lines would of course induce the flow of a current; but even without a wire — and therefore no current — a force would be induced. Dividing this force by the area enclosed by the circle gives the net electromotive force (per unit area) which "curls" around the circle. Now imagine the circle growing smaller and smaller and shrinking finally to the point *P*. By this limiting process we obtain a limiting value of the net electromotive force per unit area: this is *curl E* at *P*. Thus equation 3 says that the limiting value of electromotive force per unit area equals the rate of change of *H* at the point *P*, multiplied by the tiny negative fraction, $-\frac{1}{c}$.† Or, again, more loosely stated, a changing magnetic field creates an electric field whose electromotive force per unit area at any given point and instant of time equals the time rate of change of the magnetic field at that point and instant.

Equation 4

$$curl\, H = \frac{1}{c} \frac{\partial E}{\partial t}$$

says that, except for the change in algebraic sign (which has to do with the directions of the fields), the roles of *E* and *H* in

* In physical terms, we obtain the net capacity of the electric field to move current along the circle.

† The symbol *c*, which here stands for the ratio of the electrostatic to the electromagnetic units of electricity, is required to translate *E* (an electrostatic phenomenon) and *H* (an electromagnetic phenomenon) into the same system of units. The equation explains how Maxwell was able to connect electrical and magnetic phenomena with the velocity of light, for *c* is in fact that velocity.

equation 3 may be reversed. At any given point and instant the magnetomotive force (the analogue for magnetic fields of electromotive force) per unit of area created by a changing electric field is equal to the time rate of change of the electric field multiplied by the tiny positive fraction $\frac{1}{c}$. Now, the reader who has followed this discussion will perceive that the time rate of change of E, $\frac{\partial E}{\partial t}$, is none other than Maxwell's displacement current. For since the changes are taking place in the dielectric known as empty space, the only currents that can flow are displacement currents.* Prior to Maxwell, it was thought that the magnetic field H could be produced only by currents that flowed in wires passing through the circle. If no wires were present, the law thought to be applicable was *curl* $H = 0$. It was Maxwell's great discovery, deduced mechanically from his model and expressed mathematically in this equation, that a time-varying electric field produces (or must be accompanied by) a net "curled" magnetic force even in an insulator or empty space.†

According to Maxwell's theory, the introduction of a time-varying electric force in a dielectric produces displacement waves with the velocity of light. To put it another way, it is the surge and ebbing of the force that produces the periodic displacement waves; a static charge would merely create an instantaneous displacement, which would be fixed, but not a

* Equation 4 assumes the existence of this current and relates it quantitatively to the magnetomotive force generated by the existent magnetic field. Physically we may regard the magnetic field as creating the displacement current or, conversely, regard the displacement current as creating the accompanying magnetic field and magnetomotive force.

† Maxwell called $\frac{\partial E}{\partial t}$ the displacement current, the term "displacement" meaning that the electric field intensity E was being altered or displaced as time varies, and the term "current" suggesting that $\frac{\partial E}{\partial t}$ had the properties of a current flowing in a wire even though $\frac{\partial E}{\partial t}$ existed in empty space.

wave. Now, an electric current, as we have seen, whether in a dielectric or in a conductor, is accompanied by a magnetic force; and similarly a periodic wave of electric displacement is accompanied by a periodic magnetic force. The wave front itself, as Maxwell showed, comprises electric vibrations at right angles to the direction of propagation and a magnetic force at right angles to the electric displacement. The compound disturbance is therefore called an electromagnetic wave. A light wave (which is a displacement wave) is, as Henri Poincaré later elaborated, "a series of alternating currents, flowing in a dielectric, in the air, or in interplanetary space, changing their direction 1,000,000,000,000,000 times a second. The enormous inductive effect of these rapid alternations produces other currents in the neighboring portions of the dielectric, and thus the light waves are propagated from place to place."

The electromagnetic theory of light was testable experimentally, and indeed stood up remarkably well in laboratory trials. But this was only a limited confirmation of Maxwell's system, for if his reasoning was correct, there must be other electrical waves produced by initial disturbances of differing intensity. These waves would differ from light in wave length and would therefore not be visible, yet it should be possible to detect them with appropriate instruments. How to find them, not to say generate them, was now the crucial problem. Maxwell did not live to see it solved. Not until ten years after his death were his prophecies fulfilled and the skepticism of his most distinguished contemporaries refuted. As late as 1888 Lord Kelvin referred to Maxwell's waves as a "curious and ingenious, but not wholly tenable hypothesis"; but a year later Helmholtz's greatest pupil, Heinrich Hertz, nosed out Oliver Lodge in the race to demonstrate their existence. In a series of brilliant experiments he showed how electric waves could be "excited" (i.e., generated) by oscillation and detected by a circular conductor provided with a small gap; and how they could be polarized, reflected, refracted, made to form shadows and to interfere with each other. The connection, he said, "between light and electricity . . . of which there were hints and suspicions and even predictions in the theory, is now established. . . . Optics is no longer restricted to minute aether waves, a small fraction of a millimetre in length; its domain is extended to waves that are measured in decimetres, metres and kilometres. And in spite of this extension, it appears merely . . . as a small appendage of the great domain of electricity. We see that this latter has become a mighty kingdom."

The *Treatise*, written while Maxwell was "in retirement" at Glenlair, drew only part of his energy. As a "by-work" during the same period he wrote a textbook on heat, which appeared in 1870, and a number of papers of considerable importance on mathematics, color vision and topics of physics. He maintained a heavy scientific and social correspondence, enlarged his house, studied theology, composed stanzas of execrable verse, rode his horse, went on long walks with his dogs, visited his neighbors and played with their children, and made frequent trips to Cambridge to serve as moderator and examiner in the mathematical tripos.

In 1871 a chair in experimental physics was founded at Cambridge. It is hard to realize that at the time no courses in heat, electricity and magnetism were being taught there, and no laboratory was available for the pursuit of these arcane matters. The University, as a contemporary scholar delicately observed, "had lost touch with the great scientific movements going on outside her walls." A committee of the faculty began to bestir itself, a report was issued, and the lamentable facts fell under the gaze of the Duke of Devonshire, Chancellor of the University. He offered the money for the building and furnishing of the famous Cavendish Laboratory. Thomson, it was known, would not leave his post at Glasgow to take the new chair, and Maxwell, though at first reluctant to leave Glenlair, yielded to the urging of his friends to offer himself as a candidate. He was promptly elected.

He now devoted himself to the task of designing and superintending the erection of the laboratory. His aim was to make it the best institution of its kind, with the latest apparatus and the most effective arrangements for research. He inspected Thomson's laboratory at Glasgow and Clifton's at Oxford to learn the desirable features of both and embody them in the Cavendish. He presented to the laboratory all the apparatus in his own possession and supplemented the Duke's gift by generous money contributions. With so many details to be taken care of, the structure and its appointments were not completed until 1874. The delay, while inevitable, was inconvenient. "I have no place," wrote Maxwell, "to erect my chair, but move about like the cuckoo, depositing my notions in the Chemical Lecture Room in the first term, in the Botannical in Lent and in the Comparative Anatomy in Easter." His "notions" were the courses he gave, beginning in 1871, on heat, electricity and electromagnetism, a schedule maintained throughout the tenure of his chair. And though the audiences were often small, some of the best students were soon attracted to his lectures, which contained much important original work. The renaissance that followed in physical science at Cambridge was the direct result of his influence.

Maxwell's classic *Matter and Motion*, "a small book on a great subject," was published in 1876. About this time he contributed articles on various subjects — "Atom," "Aether," "Attraction," "Faraday," among others — to the famous ninth edition of the *Encyclopaedia Britannica*. His public lectures include a charming discourse "On the Telephone," which, though delivered when he was already very ill, is not only as clear as his best expositions but filled with gay, amusing asides. Speaking, for example, of "Professor Bell's invention," he comments on "the perfect symmetry of the whole apparatus — the wire in the middle, the two telephones at the ends of the wire, and the two gossips at the ends of the telephones. . . ." A task that occupied him for five years, almost to the very end of his life, was editing twenty packets of unpublished scientific papers of Henry Cavendish, who was great-uncle to the Duke of Devonshire. This splendid two-volume work, published in 1879, did much to fix the reputation of an immensely gifted investigator, whose important work on electricity was unknown to his contemporaries because the results were confided only to his manuscripts. Maxwell repeated Cavendish's experiments and showed that he had anticipated major discoveries in electricity, including electrostatic capacity, specific inductive capacity and Ohm's law.

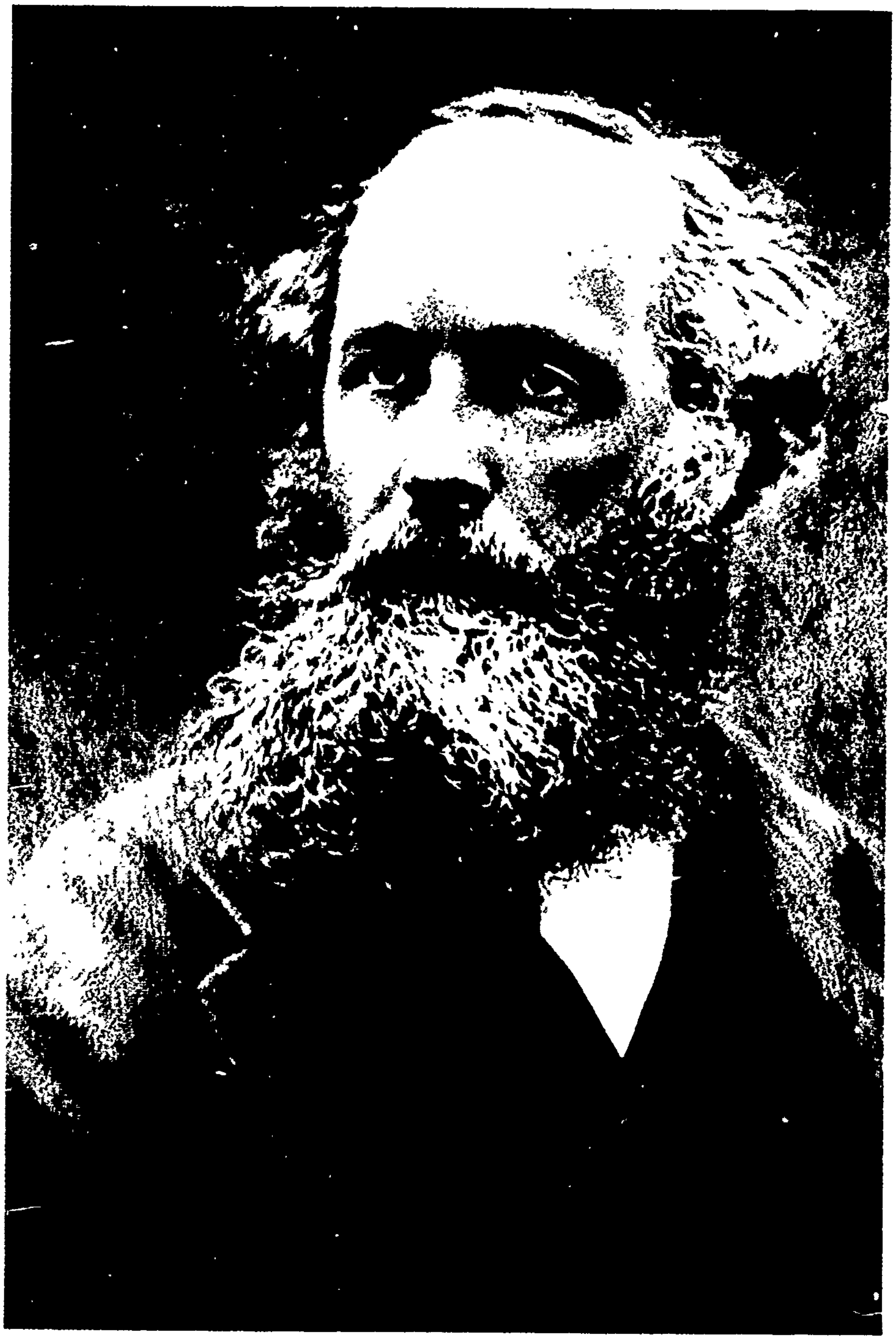
As Maxwell grew older, friends remarked on his "ever-increasing soberness" of spirit. This must not be taken to mean he was invariably melancholy or withdrawn or that his nice sense of fun — about himself no less than about others — had vanished. He continued to see his many friends, to write light verse and parodies, to promenade with his dog Toby, who was at Maxwell's side even in the laboratory, to play small practical, but never mean, jokes, to engage in what was called "humorous mystification" by advancing preposterous scientific ideas in conversation while keeping a straight face. All things, he once remarked, are "full of jokes," though they are also "quite full of solemn matters," and he was as likely to stress their light as their grave aspect.

But it is true he became somewhat more reticent with the passing years, and more and more concealed his feelings and reflections beneath an ironical shell. The tough, rational, Scotch common-sense cord of his nature had always been intertwined with threads of mysticism. Often plain, even blunt, in his address, he also had an allusive way of speaking and showed a fondness for parables. He had faith in science, yet he was at bottom skeptical as to how much could be learned from science alone about nature and meaning. It was all very well, he felt, to have "ideal aspirations"; on the other hand, "It's no use thinking of the chap ye might have been." His contemporaries remember him as both modest and intellectually scornful, tentative in his scientific opinions and dogmatic when others seemed to him to be immoderately self-assured.

"No one knows what is meant by" so-and-so was his way of answering a cocksure formulation of a scientific "truth."

The most striking of Maxwell's traits was his gentleness. "His tenderness for all living things was deep and instinctive; from earliest childhood he could not hurt a fly." An extraordinary selflessness characterized his relationship to those close to him. When his brother-in-law came to London to undergo an operation, Maxwell gave up the ground floor of his house to patient and nurse and left himself with a room so small that he frequently breakfasted on his knees because there was no room for a chair at the table. Mrs. Maxwell had a serious and prolonged illness in the last years of Maxwell's life, and he insisted on nursing her. On one occasion it is reported that he did not sleep in a bed for three weeks. But his work went on as usual, and he was as cheerful as if he enjoyed the ordeal — which may indeed have been the case. Nor did he give the slightest sign of being downcast or show self-pity when his own fatal illness seized him.

In the spring of 1877 he began to be troubled with pain and a choking sensation on swallowing. For some strange reason he consulted no one about his symptoms for almost two years, though his condition grew steadily worse. His friends at Cambridge observed that he was failing, that the spring had gone out of his step. When he went home to Glenlair for the summer of 1879, he was so obviously weakening that he called for medical help. He was in terrible pain, "hardly able to lie still for a minute together, sleepless, and with no appetite for the food which he so required." He understood thoroughly that his case was hopeless, yet his main concern seemed to be about the health of his wife. In October he was told he had only a month to live. On November 5 he died. "No man," wrote his physician, Dr. Paget, "ever met death more consciously or more calmly." When Maxwell was buried in Parton Churchyard at Glenlair, the world had not yet caught up with his ideas. Even today it has not fully explored the kingdom created by his imagination.



James Clerk Maxwell, 1831-1879.

12 Maxwell's Letters: A Collection

TRIN. COLL., Feb. 20, 1854.

DEAR THOMSON

Now that I have entered the unholy estate of bachelorhood I have begun to think of reading. This is very pleasant for some time among books of acknowledged merit wh one has not read but ought to. But we have a strong tendency to return to Physical Subjects and several of us here wish to attack Electricity.

Suppose a man to have a popular knowledge of electrical show experiments and a little antipathy to Murphy's Electricity, how ought he to proceed in reading & working so as to get a little insight into the subject wh may be of use in further reading?

If he wished to read Ampere Faraday &c how should they be arranged, and at what stage & in what order might he read your articles in the Cambridge Journal?

If you have in your mind any answer to the above questions, three of us here would be content to look upon an embodiment of it in writing as advice.

I have another question from myself. At Ardmillan while bathing on the rocks you mentioned that Gauss (?) had been investigating the bending of surfaces and had found in particular that the product of the principal radii of curvature at any point is unchanged by bending.

I have no means here of finding the paper from wh you quoted so that I w^d be obliged to you if you could give me some reference to it or even tell me whether he had considered the conditions of bending of a finite portion of a surface in general.

I have been working for some time at the more general problem & have completed the theory for surfaces of revolution and got several results i- the case of other surfaces by the consideration of two systems of lines on the surface wh may be called lines of bending.

These being given the effect of bending the surface is reduced to the consideration of one indep$_t$ var^{ble} only.

These lines themselves however are subject to certain conditions that the surface may be bent at all, and to additional conditions that these lines may continue "lines of bending".

When the lines of bending are the lines of principal curvature the conditions are very simple & are fulfilled for all surfaces of revol^{n}.

But some of the operations are long so I am going over them by a process quite different from the first.

By finding what Gauss' results are I may be spared much trouble in pruning my calculations.

Finally I have heard nothing of C. J. Taylor since he went to Glasgow, and w^{d} gladly receive information about him, Ramsay & the professorship. Are there any good classical places about Glasgow or elsewhere, "where a man might enjoy a comfortable house".

This is a letter of questions so I go on in the same spirit to the end by enquiring after the prosperity of the College especially Nos 2 & XIII i.e. commend me to the Blackburns & Mrs. Thomson.

Yrs truly

J. C. Maxwell

8 Palace Garden Terrace, Kensington, W.
1861 Dec. 10

DEAR THOMSON

I have not heard of you for some time except through Balfour Stewart who told me he had seen you lately. I hope you are now well as you are at work.

I was not farther north than Galloway last summer and we spent all our three months vacation there. Since I saw you I have been trying to develope the dynamical theory of magnetism as an affection of the whole magnetic field according to the views stated by you in the Royal Society's proceedings 1856 or Phil. Mag. 1857 vol. I p. 199 and elsewhere.

I suppose that the "magnetic medium" is divided into small portions or cells, the divisions or cell-walls being composed of a single stratum of spherical particles these particles being "electricity". The substance of the cells I suppose to be highly elastic both with respect to compression and distortion and I suppose the connexion between the cells and the particles in the cell walls to be such that there is perfect rolling without slipping between them and that they act on each other tangentially.

I then find that if the cells are set in rotation, the medium exerts a stress equivalent to a hydrostatic pressure combined with a longitudinal tension along the lines of axes of rotation.

If there be two similar systems, the first a system of magnets, electric currents and bodies capable of magnetic induction, and the second composed of cells and cell walls, the density of the cells everywhere proportional to the capacity for magnetic induction at the corresponding point of the other, and the magnitude and direction of rotation of the cells proportional to the magnetic force, then—

1. All the mechanical magnetic forces in the one system will be proportional to forces in the other arising from centrifugal force.

2. All the electric currents in the one system will be proportional to currents of the particles forming the cell walls in the other.

3. All the electromotive forces in the one system, whether arising from changes of position of magnets or currents or from motions of conductors or from changes of intensity of magnets or currents will be proportional to forces urging the particles of the cell walls arising from the tangential action of the rotating cells when their velocity is increasing or diminishing.

4. If in a non conducting body the mutual pressure of the particles of the cell walls (which corresponds to electric tension) diminishes in given direction, the particles will be urged in that direction by their mutual pressure but will be restrained by their connexion with the substance of of the cells. They will therefore produce strain in the cells till the elasticity called forth balances the tendency of the particles to move.

J. C. Maxwell to C. Hockin,

Glenlair, Dalbeattie, September 7th 1864

. . . I have been doing several electrical problems. I have got a theory of "electric absorption," i.e. residual charge, etc., and I very much want determinations of the specific induction, electric resistance, and absorption of good dielectrics, such as glass, shell-lac, gutta-percha, ebonite, sulphur, etc.

I have also cleared the electromagnetic theory of light, from all unwarrantable assumption, so that we may safely determine the velocity of light by measuring the attraction between bodies kept at a given difference of potential, the value of which is known in electromagnetic measure.

I hope there will be resistance coils at the British Association.

boundary of a solid, a line as the edge of a surface, and a point as the extremity of a line.

In like manner we may conceive the potential of a material system as a function found by a certain process of integration with respect to the masses of the bodies in the field, or we may suppose these masses themselves to have no other mathematical meaning than the volume-integrals of $\frac{1}{4\pi}\nabla^2\Psi$, where Ψ is the potential.

In electrical investigations we may use formulae in which the quantities involved are the distances of certain bodies, and the electrifications or currents in these bodies, or we may use formulae which involve other quantities, each of which is continuous through all space.

The mathematical process employed in the first method is integration along lines, over surfaces, and throughout finite spaces, those employed in the second method are partial differential equations and integrations throughout all space.

The method of Faraday seems to be intimately related to the second of these modes of treatment. He never considers bodies as existing with nothing between them but their distance, and acting on one another according to some function of that distance. He conceives all space as a field of force, the lines of force being in general curved, and those due to any body extending from it on all sides, their directions being modified by the presence of other bodies. He even speaks * of the lines of force belonging to a body as in some sense part of itself, so that in its action on distant bodies it cannot be said to act where it is not. This, however, is not a dominant idea with Faraday. I think he would rather have said that the field of space is full of lines of force, whose arrangement depends on that of the bodies in the field, and that the mechanical and electrical action on each body is determined by the lines which abut on it.

The magnetic properties of certain materials and the electric effects produced by friction were both known in ancient days. Oersted's experiment with electric current and a compass showed that electricity and magnetism are related. Maxwell found the connection between the two phenomena in his electromagnetic equations.

14 The Relationship of Electricity and Magnetism

D. K. C. MacDonald

Excerpt from his book, *Faraday, Maxwell, and Kelvin,* published in 1964.

We know that an electric current can produce forces on a magnet in its vicinity, or, in other words, an electric current produces a magnetic "field." Faraday had shown, moreover, that a changing magnetic field (produced either by moving a magnet or by varying an electric current in a coil) could induce an electric current in a neighboring, but separate, coil of wire. Thus, through these fundamental experiments of Oersted, Ampère, and particularly Faraday, various vital facts had been discovered about how electric currents and magnets could interact with one another and, as we have said earlier, these discoveries were already leading to exciting practical developments such as the electric telegraph and the submarine cables. But, in broad terms, what James Clerk Maxwell tried to do was to build up a more *general* picture of these interactions between electric and magnetic effects (or "fields") without worrying so much about actual coils of wire with electric currents in them, or about how in practice one actually produced the magnetic fields. Following Faraday's general lead in concentrating on the "lines of force" or the "fields," Maxwell tried to work out directly and quantitatively the interaction in space of the electric field on the magnetic field, and vice versa, wherever they might exist. In his mind Maxwell invented, or designed, various semi-mechanical models to build up his theory, but in the end he could discard this mental scaffolding and give a complete mathematical description of electromagnetic behavior which holds good to this day.

Consider the production of a magnetic field by a current of electricity in a coil. We know that such a current always involves a movement of electric charge, so from the electrical point of view we may say that something is changing all the time. One of the things Maxwell did was to generalize this discovery boldly, saying in essence: [I] *"A Changing Electric Field Will Always Produce a Magnetic Field."*

But, on the other hand, Faraday had shown that the movement of a magnet could produce an electric current, as we have already seen; so on the same lines this can be generalized to say: [II] *"A Changing Magnetic Field Can Produce an Electric Field."*

The ultimate result of James Clerk Maxwell's work was, in effect, that he expressed these two basic ideas in precise, quantitative terms, and he came out finally with what are now known as *Maxwell's Equations,* which, as I already have said, remain today the standard method of predicting how electricity and magnetism will behave under any given conditions. The acme of Maxwell's work, however, was his discovery that when applied in free, empty space his equations took on a form which is equally descriptive of any undamped wave motion propagating itself freely from place to place. Thus, if you drop a stone into a large pond of water a ripple or wave will proceed out from that place, and some of the energy from the falling stone will radiate outward in the wave from the splash. If you shout to somebody else some distance away, then it is a vibration or wave in the air around you which carries the sound to the distant person; or if you r[illegible]p a long, tight rope or string between two points, and then "twang" the rope, you can see a wave running along the rope, and this wave carries some of the energy that you put in the "twang." Again, if there is a violent storm at sea, the energy from this storm gets carried over long distances by waves in the ocean; the waves which smash on the rocks of Newfoundland may well be getting their energy from a storm a thousand miles or more out in the Atlantic Ocean. In each of these latter examples the waves will be damped to some degree or other. For example, waves traveling on the surface of the sea lose some energy by dragging deeper layers of water, by the very fact that water is not entirely free to move by itself, but has a viscosity or "stickiness," which means that the waves ultimately suffer losses by friction.

The particularly remarkable, and unique, feature of *electromagnetic* waves is the fact that they can propagate themselves quite freely without damping through empty space where no matter whatsoever is present, but it is not difficult to see from the two italicized statements above that a self-propelled wave motion of the electromagnetic field might be possible.

Imagine that we have electric and magnetic fields present in a small region of space, and that the fields are changing suitably with time. As the electric field changes at some point in space it will produce a magnetic field in the neighborhood, and if things are right this magnetic field will then reinforce the magnetic field in some regions, and in turn the over-all changing magnetic field will produce again a fresh electric field in its neighborhood. What Maxwell's equations showed was that this process, perhaps somewhat reminiscent of an endless game of leapfrog, could indeed be self-maintained, with the energy constantly radiating outward from where the waves started.

But this was not all. Maxwell was able to predict from this theory, moreover, the *speed* with which such an electromagnetic wave should travel in space. This speed was simply determined by the ratio of two measurements which could be made on electric and magnetic quantities in the laboratory, and it turned out that the speed predicted in this way was very close to the already known speed of light (about 300,000 km/sec $\approx$ 186,000 miles/sec). Furthermore, it is also a well-known characteristic of light that it too can propagate through empty space, as witness the light of day which reaches us unfailingly from the sun across about a hundred million miles of empty space. So Maxwell could finally say with confidence that, physically speaking, light must be a form of electromagnetic radiation.[10]

Some years after Maxwell's death, Heinrich Hertz (1857–94) was able to show experimentally, using electrical apparatus, the direct generation and detection of the electromagnetic waves predicted by Maxwell. These "Hertzian waves" are the great-grandfather of the waves which carry all our radio and television broadcasts today, and in fact radio waves, television waves, light waves, X-rays, and gamma rays, are all members of one and the same family—electromagnetic waves. In free space they all travel with identically the same speed, which for convenience we always refer to as "the velocity of light." What distinguishes one type of wave from another is simply its rate of vibration, or the corresponding wave length (i.e., the distance between two successive "crests" or "troughs" of a wave). A typical radio wave vibrates at, or has a frequency (f) of, about a million times a second ($f = 10^6$ cycles/sec = 1 M c/s), and has a wave length (λ) of about 300 meters. For those who do not mind an equation, the relationship is very simple, namely $f\lambda = c$, where c denotes, as always in physical science, the velocity of light. At the other end of the scale, a gamma ray might have a wave length of only about one ten-billionth part of a centimeter ($\lambda = 10^{-10}$ cm), and a corresponding frequency of vibration of about three hundred billion billion cycles/sec ($f = 3 \times 10^{20}$ c/s).

Electromagnetic Waves

Maxwell's electromagnetic theory also led to intense discussion later about the fundamental nature of the electromagnetic waves involved. Many physicists felt that in order to have a wave at all there had to be "something" to do the waving or vibrating, and they invented a sort of all-pervading, universal, thin soup or consommé which they called the "aether." But whether it is more reasonable to talk about electromagnetic waves in free space (which still worries some people for the same sort of reason that "action at a distance" worried people), or whether it is better to try to think about an all-permeating, vibrating "aether" is not a very burning issue today. What matters now is that Maxwell's Equations are a generally accepted foundation for discussing electromagnetic behavior under the widest range of possible situations, and also that Maxwell's lead in analyzing electromagnetism by means of the electric and magnetic fields has led more generally to the concept of discussing other forms of interaction through some appropriate "field." Indeed, Maxwell himself was at first very inclined to believe that *gravitational* attraction must also be propagated in this way, but he ran up against difficulties with the energy involved which seemed to him then insurmountable.[11]

We have seen that, starting from the picture of "action at a distance" between charges of electricity, Maxwell, following Faraday's lead, could reformulate the problem in terms of a field acting through, and at all points of, space of which the charged particles are, so to speak, now just the "terminals" or "end points." The discovery that this electromagnetic field would vibrate in free space was a great step toward identifying light as an electromagnetic wave, since the wave phenomenon of light (interference, diffraction, etc.) had been known for a long time. At the same time there had always been some persistent reasons for regarding light alternatively as a corpuscular phenomenon, and Einstein was to show, half a century later, that Maxwell's vibrating electromagnetic aether, when coupled with Planck's quantum theory first proposed around 1900, could also then be regarded in a more or less corpuscular manner. What Planck and Einstein showed was that the energy in the electromagnetic field could only exist in certain minimum-sized bundles or "quanta" dependent in magnitude on the frequency of vibration and the newly discovered Planck's constant. These "bundles" of light, or more technically "quanta" of the electromagnetic field, are generally known today as photons. So now we can think of electromagnetic interactions as either conveyed by the vibrating aether or equivalently as conveyed by streams of photons which will to some extent behave like particles. In dealing with many kinds of interactions, including those which hold an atomic nucleus together, modern physics finds it most valuable to be able to think in both these

...displacement on the supposition ...ubic and linear elasticities are connected as in a "per- ...ound from this the attraction between two bodies having given quantities of free electricity on their surfaces, and then by comparison with Weber's value of the statical measure of a unit of electrical current I have deduced the relation between the elasticity and density of the cells. The velocity of transverse undulations follows from this directly and is equal to 193088 miles per second, very nearly that of light.

Velocity of light miles per sec.	192,500 by aberration
	195,777 by Fizeau
	193,118 Galbraith & Haughton's statement of Fizeau's results

I made out the equations in the country before I had any suspicion of the nearness between the two values of the velocity of propagation of magnetic effects and that of light, so that I think I have reason to believe that the magnetic and luminiferous media are identical and that Weber's number is really, as it appears to be, one half the velocity of light in millimetres per second.

If there is in all media, in spite of the disturbing influence of gross matter, the same relation between the velocity of light and the statical action of electricity, then the "dielectric capacity", that is, the capacity of a Leyden jar of given thickness formed of it, is proportional to the square of the index of refraction.*

Do you know any good measures of dielectric capacity of transparent substances? I have read Faraday & Harris on the subject and I think they are likely to be generally too small. I think Fleeming Jenkin has found that of gutta percha caoutchouc &c. Where can one find his method, and what method do you recommend?

I think I see a way to work with flat plates of different substances along with your divided ring. electrometer.

A & E are plates connected with each other, and with a source of electricity; C is connected with the ground, B & D are moveable plates connected with the two halves of the ring. F is a dielectric. Then by a proper placing of D things may be arranged so that the electrification of A and E produces no difference of potentials in the divided ring, after which the capacity of F follows by calculation.

terms without being bound to regard one picture as more necessarily "real" than the other.

the rotation of the plane of polarization in the same direction as the angular momentum of all the vortices, the rotation being proportional to

A the thickness of the medium
B the magnetic intensity along the axis
C the index of refraction in the medium
D inversely as the square of the wave length in air
F directly as the radius of the vortices
G " the magnetic capacity.

I have been seeking for experiments lately made which I have lost sight of showing that the rotation varies faster than the inverse square of wave length in air, so that C & D give the true law.

A & B are proved by Verdet,

F is not yet capable of proof

G is consistent with Verdet's result that the rotation in salts of iron is opposite to that in diamagnetic substances. I think that molecules of iron are set in motion by the cells and revolve the opposite way so as to produce a very great energy of rotation but an angular momentum in the opposite direction to that of the vortices.

I find that unless the diameter of the vortices is sensible, no result is likely to be obtained by making a magnet revolve freely about an axis perp. to the magnetic axis. I have tried it but have not yet got rid of the effects of terrestrial magnetism which are very strong on a powerful electromagnet which I use. I think it more probable that a coil of conducting wire might be found to have a slight shock tending to turn it about its axis when the electricity is let on or cut off, or that a piece of iron magnetized by a helix might have a similar impulse.

Yours truly

J. C. MAXWELL

The formulation of Maxwell's equations opened the new area of science called electromagnetism, with far-reaching consequences.

15 The Electromagnetic Field

Albert Einstein and Leopold Infeld

Excerpt from their book entitled the *Evolution of Physics* published in 1938 and 1961.

DURING the second half of the nineteenth century new and revolutionary ideas were introduced into physics; they opened the way to a new philosophical view, differing from the mechanical one. The results of the work of Faraday, Maxwell, and Hertz led to the development of modern physics, to the creation of new con-

DEAR STOKES,

I have been reading Jamin's note on the Theory of Reflexion and Refraction, Ann. de Ch. 1860, pt. i. p. 413.

I am not yet able to satisfy myself about the conditions to be fulfilled at the surface except of course the condition of conservation of energy.

Jamin insists on the equality of the motion both horizontal and vertical in the two media. I do not see the necessity for equality of motion; but I think action and reaction must be equal between the media, provided the media pure and simple vibrate and nothing along with them.

If the gross matter in each medium does not vibrate, or has a different phase and amplitude from the ether, then there will be six relations between the four quantities:—two portions of ether and two kinds of gross matter.

Have you written anything about the rival theories of reflexion? or can you tell me of any thing you agree with or eminently differ from on that subject? I think you once told me that the subject was a stiff one to the best skilled in undulations.

Jamin deduces (p. 422) from his conditions of equality of motion in the two media for vibrations in the plane of incidence that the density of the medium is the same in all substances.

That is to say he gets this by pure mathematics without any experiment.

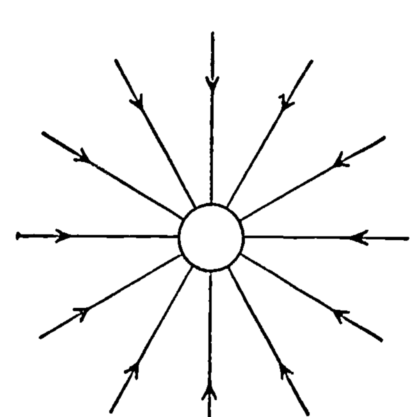
Or according to him no such vibrations could exist in the media unless they were of equal density.

This I think simply disproves his original assumption of the equality of the displacements in the two media.

In fact the equality of displacements combined with the equality of energy involves the equality of density.

Therefore the general theory, which ought to be able to explain the case of media of unequal density (even if there were none such) must not assume equality of displacements of contiguous particles on each side of the surface.

We can represent this fact in a new way, and shall do so even though it is difficult to understand the advan-



tage of this. The small circle in our drawing represents an attracting body, say, the sun. Actually, our diagram should be imagined as a model in space and not as a drawing on a plane. Our small circle, then, stands for a sphere in space, say, the sun. A body, the so-called *test body*, brought somewhere within the vicinity of the sun will be attracted along the line connecting the centers of the two bodies. Thus the lines in our draw-

But there is nothing in the surface of separation of two media analogous to this gluing together that I can detect.

I have now got materials for calculating the velocity of transmission of a magnetic disturbance through air founded on experimental evidence, without any hypothesis about the structure of the medium or any mechanical explanation of electricity or magnetism.

The result is that only transverse disturbances can be propagated and that the velocity is that found by Weber and Kohlrausch which is nearly that of light. This is the velocity with which such slow disturbances as we can make would be propagated. If the same law holds for rapid ones, then there is no difference between polarized light and rapid electromagnetic disturbances in one plane.

I have written out so much of the theory as does not involve the conditions at bounding surfaces, and will send it to the R. S. in a week.

I am trying to understand the conditions at a surface for reflexion and refraction, but they may not be the same for the period of vibration of light and for experiments made at leisure.

We are devising methods to determine this velocity = electromagnetic : electrostatic unit of electricity. Thomson is going to weigh an electromotive force. Jenkin and I are going to measure the capacity of a conductor both ways and I have a plan of direct equilibrium between an electromagnetic repulsion and electrostatic attraction.

Yours truly,

J. C. MAXWELL

a test body would behave if brought into the vicinity of the sphere for which the field is constructed.

The lines in our space model are always perpendicular to the surface of the sphere. Since they diverge from one point, they are dense near the sphere and become less and less so farther away. If we increase the distance from the sphere twice or three times, then the density of the lines, in our space-model, though not in the drawing, will be four or nine times less. Thus the lines serve a double purpose. On the one hand they show the direction of the force acting on a body brought into the neighborhood of the sphere-sun. On the other hand the density of the lines of force in space shows how the force varies with the distance. The drawing of the field, correctly interpreted, represents the direction of the gravitational force and its dependence on distance. One can read the law of gravitation from such a drawing just as well as from a description of the action in words, or in the precise and economical language of mathematics. This *field representation*, as we shall call it, may appear clear and interesting but there is no reason to believe that it marks any real adreasonable person, an attempt to make our drawing something more than a model leads nowhere.

We do not intend, however, to discuss the gravitational problem just now. It served only as an introduction, simplifying the explanation of similar methods of reasoning in the theory of electricity.

We shall begin with a discussion of the experiment which created serious difficulties in our mechanical interpretation. We had a current flowing through a wire circuit in the form of a circle. In the middle of the circuit was a magnetic needle. The moment the current began to flow a new force appeared, acting on the magnetic pole, and perpendicular to any line connecting the wire and the pole. This force, if caused by a circulating charge, depended, as shown by Rowland's experiment, on the velocity of the charge. These experimental facts contradicted the philosophical view that all forces must act on the line connecting the particles and can depend only upon distance.

The exact expression for the force of a current acting on a magnetic pole is quite complicated, much more so, indeed, than the expression for gravitational forces. We can, however, attempt to visualize the actions just as we did in the case of a gravitational force. Our question is: with what force does the current act upon a magnetic pole placed somewhere in its vicinity? It would be rather difficult to describe this force in words. Even a mathematical formula would be complicated and awkward. It is best to represent all we know about the acting forces by a drawing, or rather by a spatial model, with lines of force. Some difficulty is caused by the fact that a magnetic pole exists only in connection with another magnetic pole, form-

From his *Treatise on Electricity and Magnetism* published in 1873.

528.] THE discovery by Örsted of the magnetic action of an electric current led by a direct process of reasoning to that of magnetization by electric currents, and of the mechanical action between electric currents. It was not, however, till 1831 that Faraday, who had been for some time endeavouring to produce electric currents by magnetic or electric action, discovered the conditions of magneto-electric induction. The method which Faraday employed in his researches consisted in a constant appeal to experiment as a means of testing the truth of his ideas, and a constant cultivation of ideas under the direct influence of experiment. In his published researches we find these ideas expressed in language which is all the better fitted for a nascent science, because it is somewhat alien from the style of physicists who have been accustomed to established mathematical forms of thought.

The experimental investigation by which Ampère established the laws of the mechanical action between electric currents is one of the most brilliant achievements in science.

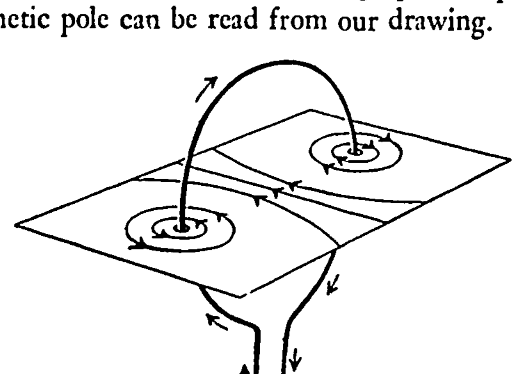
The whole, theory and experiment, seems as if it had leaped, full grown and full armed, from the brain of the 'Newton of electricity.' It is perfect in form, and unassailable in accuracy, and it is summed up in a formula from which all the phenomena may be deduced, and which must always remain the cardinal formula of electro-dynamics.

The method of Ampère, however, though cast into an inductive form, does not allow us to trace the formation of the ideas which guided it. We can scarcely believe that Ampère really discovered the law of action by means of the experiments which he describes. We are led to suspect, what, indeed, he tells us himself*, that he

* *Théorie des Phenomènes Electrodynamiques*, p. 9.

ing a dipole. We can, however, always imagine the magnetic needle of such length that only the force acting upon the pole nearer the current has to be taken into account. The other pole is far enough away for the force acting upon it to be negligible. To avoid ambiguity we shall say that the magnetic pole brought nearer to the wire is the *positive* one.

The character of the force acting upon the positive magnetic pole can be read from our drawing.



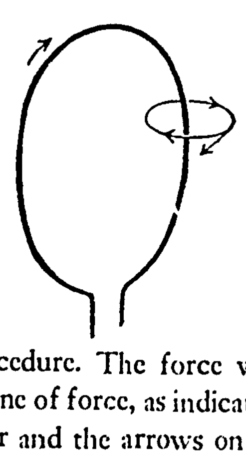
First we notice an arrow near the wire indicating the direction of the current, from higher to lower potential. All other lines are just lines of force belonging to coverer. Every student therefore should read Ampère's research as a splendid example of scientific style in the statement of a discovery, but he should also study Faraday for the cultivation of a scientific spirit, by means of the action and reaction which will take place between the newly discovered facts as introduced to him by Faraday and the nascent ideas in his own mind.

It was perhaps for the advantage of science that Faraday, though thoroughly conscious of the fundamental forms of space, time, and force, was not a professed mathematician. He was not tempted to enter into the many interesting researches in pure mathematics which his discoveries would have suggested if they had been exhibited in a mathematical form, and he did not feel called upon either to force his results into a shape acceptable to the mathematical taste of the time, or to express them in a form which mathematicians might attack. He was thus left at leisure to do his proper work, to coordinate his ideas with his facts, and to express them in natural, untechnical language.

It is mainly with the hope of making these ideas the basis of a mathematical method that I have undertaken this treatise.

529.] We are accustomed to consider the universe as made up of parts, and mathematicians usually begin by considering a single particle, and then conceiving its relation to another particle, and so on. This has generally been supposed the most natural method. To conceive of a particle, however, requires a process of abstraction, since all our perceptions are related to extended bodies, so that the idea of the *all* that is in our consciousness at a given instant is perhaps as primitive an idea as that of any individual thing. Hence there may be a mathematical method in which we proceed from the whole to the parts instead of from the parts to the whole. For example, Euclid, in his first book, conceives a line as traced out by a point, a surface as swept out by a line, and a solid as generated by a surface. But he also defines a surface as the

such a model is not as simple as in our previous example, where the lines of force were straight. In our next diagram only one line of force is drawn in order to

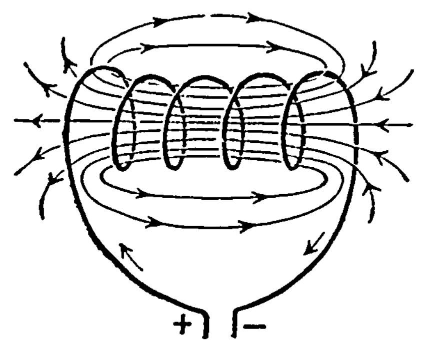


clarify the procedure. The force vector lies on the tangent to the line of force, as indicated. The arrow of the force vector and the arrows on the line of force point in the same direction. Thus this is the direction in which the force acts on a magnetic pole at this point. A good drawing, or rather a good model, also tells us something about the length of the force vector at any point. This vector has to be longer where the lines are denser, i.e., near the wire, shorter where the lines are less dense, i.e., far from the wire.

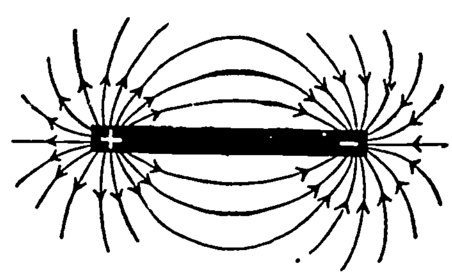
we come once more to the conclusion that the force acts in a direction perpendicular to any line connecting the wire and the pole, for the tangent to a circle is always perpendicular to its radius. Our entire knowledge of the acting forces can be summarized in the construction of the field. We sandwich the concept of the field between that of the current and that of the magnetic pole in order to represent the acting forces in a simple way.

Every current is associated with a magnetic field, i.e., a force always acts on a magnetic pole brought near the wire through which a current flows. We may remark in passing that this property enables us to construct sensitive apparatus for detecting the existence of a current. Once having learned how to read the character of the magnetic forces from the field model of a current, we shall always draw the field surrounding the wire through which the current flows, in order to represent the action of the magnetic forces at any point in space. Our first example is the so-called solenoid. This is, in fact, a coil of wire as shown in the drawing. Our aim is to learn, by experiment, all we can about the magnetic field associated with the current flowing through a solenoid and to incorporate this knowledge in the construction of a field. A drawing represents our result. The curved lines of force are closed, and surround the solenoid in a way characteristic of the magnetic field of a current. See top p. 132.

The field of a bar magnet can be represented in the same way as that of a current. Another drawing shows this. The lines of force are directed from the positive to the negative pole. The force vector always lies on



the tangent to the line of force and is longest near the poles because the density of the lines is greatest at these points. The force vector represents the action of the magnet on a positive magnetic pole. In this case the magnet and not the current is the "source" of the field.



Our last two drawings should be carefully compared. In the first, we have the magnetic field of a current flowing through a solenoid; in the second, the field of a bar magnet. Let us ignore both the solenoid and the bar and observe only the two outside fields. We immediately notice that they are of exactly the same character; in each case the lines of force lead from one end of the solenoid or bar to the other.

The field representation yields its first fruit! It would be rather difficult to see any strong similarity between the current flowing through a solenoid and a bar magnet if this were not revealed by our construction of the field.

The concept of field can now be put to a much more severe test. We shall soon see whether it is anything more than a new representation of the acting forces. We could reason: assume, for a moment, that the field characterizes all actions determined by its sources in a unique way. This is only a guess. It would mean that if a solenoid and a bar magnet have the same field, then all their influences must also be the same. It would mean that two solenoids, carrying electric currents, behave like two bar magnets, that they attract or repel each other depending, exactly as in the case of bars, on their relative positions. It would also mean that a solenoid and a bar attract or repel each other in the same way as two bars. Briefly speaking, it would mean that all actions of a solenoid through which a current flows, and of a corresponding bar magnet are the same, since the field alone is responsible for them, and the field in both cases is of the same character. Experiment fully confirms our guess!

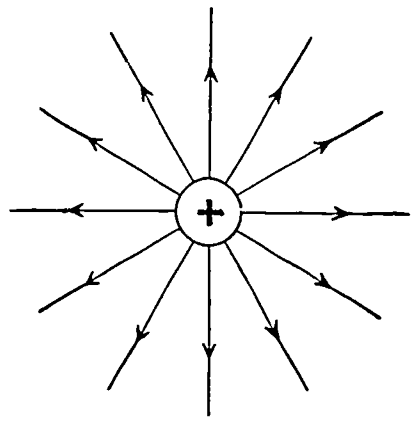
How difficult it would be to find those facts without the concept of field! The expression for a force acting between a wire through which a current flows and a magnetic pole is very complicated. In the case of two solenoids we should have to investigate the forces with which two currents act upon each other. But if we do this, with the help of the field, we immediately notice the character of all those actions at the moment when the similarity between the field of a solenoid and that of a bar magnet is seen.

We have the right to regard the field as something much more than we did at first. The properties of the field alone appear to be essential for the description of phenomena; the differences in source do not matter. The concept of field reveals its importance by leading to new experimental facts.

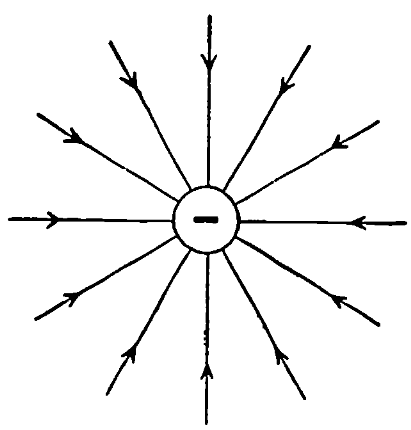
The field proved a very helpful concept. It began as something placed between the source and the magnetic needle in order to describe the acting force. It was thought of as an "agent" of the current, through which all action of the current was performed. But now the agent also acts as an interpreter, one who translates the laws into a simple, clear language, easily understood.

The first success of the field description suggests that it may be convenient to consider all actions of currents, magnets and charges indirectly, i.e., with the help of the field as an interpreter. A field may be regarded as something always associated with a current. It is there even in the absence of a magnetic pole to test its existence. Let us try to follow this new clew consistently.

The field of a charged conductor can be introduced in much the same way as the gravitational field, or the field of a current or magnet. Again only the simplest example! To design the field of a positively charged sphere, we must ask what kind of forces are acting on a small positively charged test body brought near the source of the field, the charged sphere. The fact that we use a positively and not a negatively charged test body is merely a convention, indicating in which direction the arrows on the line of force should be drawn. The model is analogous to that of a gravitational field (p. 126) because of the similarity between Coulomb's law and Newton's. The only difference between the



two models is that the arrows point in opposite directions. Indeed, we have repulsion of two positive charges and attraction of two masses. However, the field of a sphere with a negative charge will be identical with a gravitational field since the small positive testing charge will be attracted by the source of the field.



If both electric and magnetic poles are at rest, there is no action between them, neither attraction nor repulsion. Expressing the same fact in the field language we can say: an electrostatic field does not influence a magnetostatic one and vice versa. The words "static field" mean a field that does not change with time. The magnets and charges would rest near one another for an eternity if no external forces disturbed them. Electrostatic, magnetostatic and gravitational fields are all of different character. They do not mix; each preserves its individuality regardless of the others.

Let us return to the electric sphere which was, until now, at rest, and assume that it begins to move due to the action of some external force. The charged sphere moves. In the field language this sentence reads: the field of the electric charge changes with time. But the motion of this charged sphere is, as we already know from Rowland's experiment, equivalent to a current. Further, every current is accompanied by a magnetic field. Thus the chain of our argument is:

$$\begin{array}{ccl} \text{motion of charge} & \rightarrow & \text{change of an electric field} \\ \downarrow & & \\ \text{current} & \rightarrow & \text{associated magnetic field.} \end{array}$$

We, therefore, conclude: *The change of an electric field produced by the motion of a charge is always accompanied by a magnetic field.*

Our conclusion is based on Oersted's experiment but it covers much more. It contains the recognition that the association of an electric field, changing in time, with a magnetic field is essential for our further argument.

As long as a charge is at rest there is only an electrostatic field. But a magnetic field appears as soon as the charge begins to move. We can say more. The magnetic field created by the motion of the charge will be stronger if the charge is greater and if it moves faster. This also is a consequence of Rowland's experiment. Once again using the field language, we can say: the faster the electric field changes, the stronger the accompanying magnetic field.

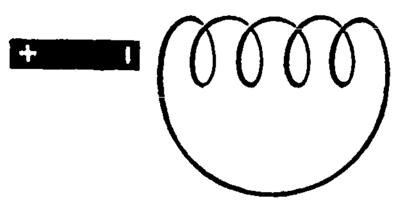
We have tried here to translate familiar facts from the language of fluids, constructed according to the old mechanical view, into the new language of fields. We shall see later how clear, instructive, and far-reaching our new language is.

THE TWO PILLARS OF THE FIELD THEORY

"The change of an electric field is accompanied by a magnetic field." If we interchange the words "magnetic" and "electric," our sentence reads: "The change of a magnetic field is accompanied by an electric field." Only an experiment can decide whether or not this statement is true. But the idea of formulating this problem is suggested by the use of the field language.

Just over a hundred years ago, Faraday performed an experiment which led to the great discovery of induced currents.

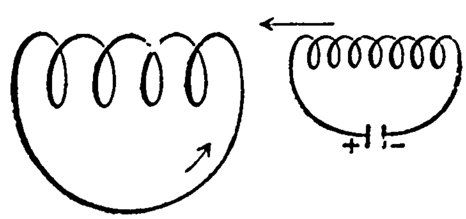
The demonstration is very simple. We need only a solenoid or some other circuit, a bar magnet, and one of the many types of apparatus for detecting the existence of an electric current. To begin with, a bar magnet is kept at rest near a solenoid which forms a closed circuit. No current flows through the wire, for no source is present. There is only the magnetostatic field of the bar magnet which does not change with time. Now, we quickly change the position of the magnet either by removing it or by bringing it nearer the solenoid, whichever we prefer. At this moment, a current will appear for a very short time and then vanish.



Whenever the position of the magnet is changed, the current reappears, and can be detected by a sufficiently sensitive apparatus. But a current—from the point of view of the field theory—means the existence of an electric field forcing the flow of the electric fluids through the wire. The current, and therefore the electric field, too, vanishes when the magnet is again at rest.

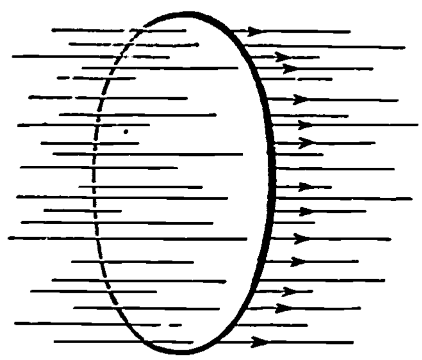
Imagine for a moment that the field language is unknown and the results of this experiment have to be described, qualitatively and quantitatively, in the language of old mechanical concepts. Our experiment then shows: by the motion of a magnetic dipole a new force was created, moving the electric fluid in the wire. The next question would be: upon what does this force depend? This would be very difficult to answer. We should have to investigate the dependence of the force upon the velocity of the magnet, upon its shape, and upon the shape of the circuit. Furthermore, this experiment, if interpreted in the old language, gives us no hint at all as to whether an induced current can be excited by the motion of another circuit carrying a current, instead of by motion of a bar magnet.

It is quite a different matter if we use the field language and again trust our principle that the action is determined by the field. We see at once that a solenoid through which a current flows would serve as well as a bar magnet. The drawing shows two solenoids: one, small, through which a current flows, and the other, in which the induced current is detected, larger. We



could move the small solenoid, as we previously moved the bar magnet, creating an induced current in the larger solenoid. Furthermore, instead of moving the small solenoid, we could create and destroy a magnetic field by creating and destroying the current, that is, by opening and closing the circuit. Once again, new facts suggested by the field theory are confirmed by experiment!

Let us take a simpler example. We have a closed wire without any source of current. Somewhere in the vicinity is a magnetic field. It means nothing to us whether the source of this magnetic field is another circuit through which an electric current flows, or a bar magnet. Page 140 shows the closed circuit and the magnetic lines of force. The qualitative and quantitative description of the induction phenomena is very simple in terms of the field language. As marked on the drawing, some lines of force go through the surface bounded by the wire. We have to consider the lines of force cutting that part of the plane which has the wire



for a rim. No electric current is present so long as the field does not change, no matter how great its strength. But a current begins to flow through the rim-wire as soon as the number of lines passing through the surface surrounded by wire changes. The current is determined by the change, however it may be caused, of the number of lines passing the surface. This change in the number of lines of force is the only essential concept for both the qualitative and the quantitative descriptions of the induced current. "The number of lines changes" means that the density of the lines changes and this, we remember, means that the field strength changes.

These then are the essential points in our chain of reasoning: change of magnetic field → induced current → motion of charge → existence of an electric field.

Therefore: *a changing magnetic field is accompanied by an electric field.*

Thus we have found the two most important pillars of support for the theory of the electric and magnetic field. The first is the connection between the changing electric field and the magnetic field. It arose from Oersted's experiment on the deflection of a magnetic needle and led to the conclusion: *a changing electric field is accompanied by a magnetic field.*

The second connects the changing magnetic field with the induced current and arose from Faraday's experiment. Both formed a basis for quantitative description.

Again the electric field accompanying the changing magnetic field appears as something real. We had to imagine, previously, the magnetic field of a current existing without the testing pole. Similarly, we must claim here that the electric field exists without the wire testing the presence of an induced current.

In fact, our two-pillar structure could be reduced to only one, namely, to that based on Oersted's experiment. The result of Faraday's experiment could be deduced from this with the law of conservation of energy. We used the two-pillar structure only for the sake of clearness and economy.

One more consequence of the field description should be mentioned. There is a circuit through which a current flows, with for instance, a voltaic battery as the source of the current. The connection between the wire and the source of the current is suddenly broken. There is, of course, no current now! But during this short interruption an intricate process takes place, a process which could again have been foreseen by the field theory. Before the interruption of the current there was a magnetic field surrounding the wire. This ceased to exist the moment the current was interrupted. Therefore, through the interruption of a current, a magnetic field disappeared. The number of lines of force passing through the surface surrounded by the wire changed very rapidly. But such a rapid change, however it is produced, must create an induced current. What really matters is the change of the magnetic field making the induced current stronger if the change is greater. This consequence is another test for the theory. The disconnection of a current must be accompanied by the appearance of a strong, momentary induced current. Experiment again confirms the prediction. Anyone who has ever disconnected a current must have noticed that a spark appears. This spark reveals the strong potential differences caused by the rapid change of the magnetic field.

The same process can be looked at from a different point of view, that of energy. A magnetic field disappeared and a spark was created. A spark represents energy, therefore, so also must the magnetic field. To use the field concept and its language consistently, we must regard the magnetic field as a store of energy. Only in this way shall we be able to describe the electric and magnetic phenomena in accordance with the law of conservation of energy.

Starting as a helpful model the field became more and more real. It helped us to understand old facts and led us to new ones. The attribution of energy to the field is one step further in the development in which the field concept was stressed more and more, and the concepts of substances, so essential to the mechanical point of view, were more and more suppressed.

THE REALITY OF THE FIELD

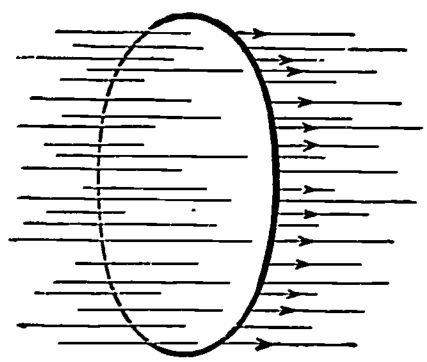
The quantitative, mathematical description of the laws of the field is summed up in what are called Maxwell's equations. The facts mentioned so far led to the formulation of these equations but their content is much richer than we have been able to indicate. Their simple form conceals a depth revealed only by careful study.

The formulation of these equations is the most important event in physics since Newton's time, not only because of their wealth of content, but also because they form a pattern for a new type of law.

The characteristic features of Maxwell's equations, appearing in all other equations of modern physics, are summarized in one sentence. Maxwell's equations are laws representing the *structure* of the field.

Why do Maxwell's equations differ in form and character from the equations of classical mechanics? What does it mean that these equations describe the structure of the field? How is it possible that, from the results of Oersted's and Faraday's experiments, we can form a new type of law, which proves so important for the further development of physics?

We have already seen, from Oersted's experiment, how a magnetic field coils itself around a changing electric field. We have seen, from Faraday's experiment, how an electric field coils itself around a changing magnetic field. To outline some of the characteristic features of Maxwell's theory, let us, for the moment, focus all our attention on one of these experiments, say, on that of Faraday. We repeat the drawing in which an electric current is induced by a changing magnetic field. We already know that an induced current appears if the number of lines of force, passing the surface bounded by the wire, changes. Then the current



will appear if the magnetic field changes or the circuit is deformed or moved: if the number of magnetic lines passing through the surface is changed, no matter how this change is caused. To take into account all these various possibilities, to discuss their particular influences, would necessarily lead to a very complicated theory. But can we not simplify our problem? Let us try to eliminate from our considerations everything which refers to the shape of the circuit, to its length, to the surface enclosed by the wire. Let us imagine that the circuit in our last drawing becomes smaller and smaller, shrinking gradually to a very small circuit enclosing a certain point in space. Then everything concerning shape and size is quite irrelevant. In this limiting process where the closed curve shrinks to a point, size and shape automatically vanish from our considerations and we obtain laws connecting changes of magnetic and electric field at an arbitrary point in space at an arbitrary instant.

Thus, this is one of the principal steps leading to Maxwell's equations. It is again an idealized experiment performed in imagination by repeating Faraday's experiment with a circuit shrinking to a point.

We should really call it half a step rather than a whole one. So far our attention has been focused on Faraday's experiment. But the other pillar of the field theory, based on Oersted's experiment, must be considered just as carefully and in a similar manner. In this experiment the magnetic lines of force coil themselves around the current. By shrinking the circular magnetic lines of force to a point, the second half-step is performed and the whole step yields a connection between the changes of the magnetic and electric fields at an arbitrary point in space and at an arbitrary instant.

But still another essential step is necessary. According to Faraday's experiment, there must be a wire testing the existence of the electric field, just as there must be a magnetic pole, or needle, testing the existence of a magnetic field in Oersted's experiment. But Maxwell's new theoretical idea goes beyond these experimental facts. The electric and magnetic field, or in short, the *electromagnetic* field is, in Maxwell's theory, something real. The electric field is produced by a changing magnetic field, quite independently, whether or not there is a wire to test its existence; a magnetic field is produced by a changing electric field, whether or not there is a magnetic pole to test its existence.

Thus two essential steps led to Maxwell's equations. The first: in considering Oersted's and Rowland's experiments, the circular line of the magnetic field coiling itself around the current and the changing electric field, had to be shrunk to a point; in considering Faraday's experiment, the circular line of the electric field coiling itself around the changing magnetic field had to be shrunk to a point. The second step consists of the realization of the field as something real; the electromagnetic field once created exists, acts, and changes according to Maxwell's laws.

Maxwell's equations describe the structure of the electromagnetic field. All space is the scene of these laws and not, as for mechanical laws, only points in which matter or charges are present.

We remember how it was in mechanics. By knowing the position and velocity of a particle at one single instant, by knowing the acting forces, the whole future path of the particle could be forseen. In Maxwell's theory, if we know the field at one instant only, we can deduce from the equations of the theory how the whole field will change in space and time. Maxwell's equations enable us to follow the history of the field, just as the mechanical equations enabled us to follow the history of material particles.

But there is still one essential difference between mechanical laws and Maxwell's laws. A comparison of Newton's gravitational laws and Maxwell's field laws will emphasize some of the characteristic features expressed by these equations.

With the help of Newton's laws we can deduce the motion of the earth from the force acting between the sun and the earth. The laws connect the motion of the earth with the action of the far-off sun. The earth and the sun, though so far apart, are both actors in the play of forces.

In Maxwell's theory there are no material actors. The mathematical equations of this theory express the laws governing the electromagnetic field. They do not, as in Newton's laws, connect two widely separated events; they do not connect the happenings *here* with the conditions *there*. The field *here* and *now* depends on the field in the *immediate neighborhood* at a time *just past*. The equations allow us to predict what will happen a little further in space and a little later in time, if we know what happens here and now. They allow us to increase our knowledge of the field by small steps. We can deduce what happens here from that which happened far away by the summation of these very small steps. In Newton's theory, on the contrary, only big steps connecting distant events are permissible. The experiments of Oersted and Faraday can be regained from Maxwell's theory, but only by the summation of small steps each of which is governed by Maxwell's equations.

A more thorough mathematical study of Maxwell's equations shows that new and really unexpected conclusions can be drawn and the whole theory submitted to a test on a much higher level, because the theoretical consequences are now of a quantitative character and are revealed by a whole chain of logical arguments.

Let us again imagine an idealized experiment. A small sphere with an electric charge is forced, by some external influence, to oscillate rapidly and in a rhythmical way, like a pendulum. With the knowledge we already have of the changes of the field, how shall we describe everything that is going on here, in the field language?

The oscillation of the charge produces a changing electric field. This is always accompanied by a changing magnetic field. If a wire forming a closed circuit is placed in the vicinity, then again the changing magnetic field will be accompanied by an electric current in the circuit. All this is merely a repetition of known facts, but the study of Maxwell's equations gives a much deeper insight into the problem of the oscillating electric charge. By mathematical deduction from Maxwell's equations we can detect the character of the field surrounding an oscillating charge its structure near and far from the source and its change with time. The outcome of such deduction is the *electromagnetic wave*. Energy radiates from the oscillating charge traveling with a definite speed through space; but a transference of energy, the motion of a state, is characteristic of all wave phenomena.

Different types of waves have already been considered. There was the longitudinal wave caused by the pulsating sphere, where the changes of density were propagated through the medium. There was the jellylike medium in which the transverse wave spread. A deformation of the jelly, caused by the rotation of the sphere, moved through the medium. What kind of changes are now spreading in the case of an electromagnetic wave? Just the changes of an electromagnetic field! Every change of an electric field produces a magnetic field; every change of this magnetic field produces an electric field; every change of . . . , and so on. As field represents energy, all these changes spreading out in space, with a definite velocity, produce a wave. The electric and magnetic lines of force always lie, as deduced from the theory, on planes perpendicular to the direction of propagation. The wave produced is, therefore, transverse. The original features of the picture of the field we formed from Oersted's and Faraday's experiments are still preserved, but we now recognize that it has a deeper meaning.

The electromagnetic wave spreads in empty space.

This, again, is a consequence of the theory. If the oscillating charge suddenly ceases to move, then, its field becomes electrostatic. But the series of waves created by the oscillation continues to spread. The waves lead an independent existence and the history of their changes can be followed just as that of any other material object.

We understand that our picture of an electromagnetic wave, spreading with a certain velocity in space and changing in time, follows from Maxwell's equations only because they describe the structure of the electromagnetic field at any point in space and for any instant.

There is another very important question. With what speed does the electromagnetic wave spread in empty space? The theory, with the support of some data from simple experiments having nothing to do with the actual propagation of waves, gives a clear answer: *the velocity of an electromagnetic wave is equal to the velocity of light.*

Oersted's and Faraday's experiments formed the basis on which Maxwell's laws were built. All our results so far have come from a careful study of these laws, expressed in the field language. The theoretical discovery of an electromagnetic wave spreading with the speed of light is one of the greatest achievements in the history of science.

Experiment has confirmed the prediction of theory. Fifty years ago, Hertz proved, for the first time, the existence of electromagnetic waves and confirmed experimentally that their velocity is equal to that of light. Nowadays, millions of people demonstrate that electromagnetic waves are sent and received. Their apparatus is far more complicated than that used by Hertz and detects the presence of waves thousands of miles from their sources instead of only a few yards.

16

Radiation Belts Around the Earth

James Van Allen

1959

So far, the most interesting and least expected result of man's exploration of the immediate vicinity of the earth is the discovery that our planet is ringed by a region—to be exact, two regions—of high-energy radiation extending many thousands of miles into space. The discovery is of course troubling to astronauts; somehow the human body will have to be shielded from this radiation, even on a rapid transit through the region. But geophysicists, astrophysicists, solar astronomers and cosmic-ray physicists are enthralled by the fresh implications of these findings. The configuration of the region and the radiation it contains bespeak a major physical phenomenon involving cosmic rays and solar corpuscles in the vicinity of the earth. This enormous reservoir of charged particles plays a still-unexplained role as middleman in the interaction of earth and sun which is reflected in magnetic storms, in the airglow and in the beautiful displays of the aurora.

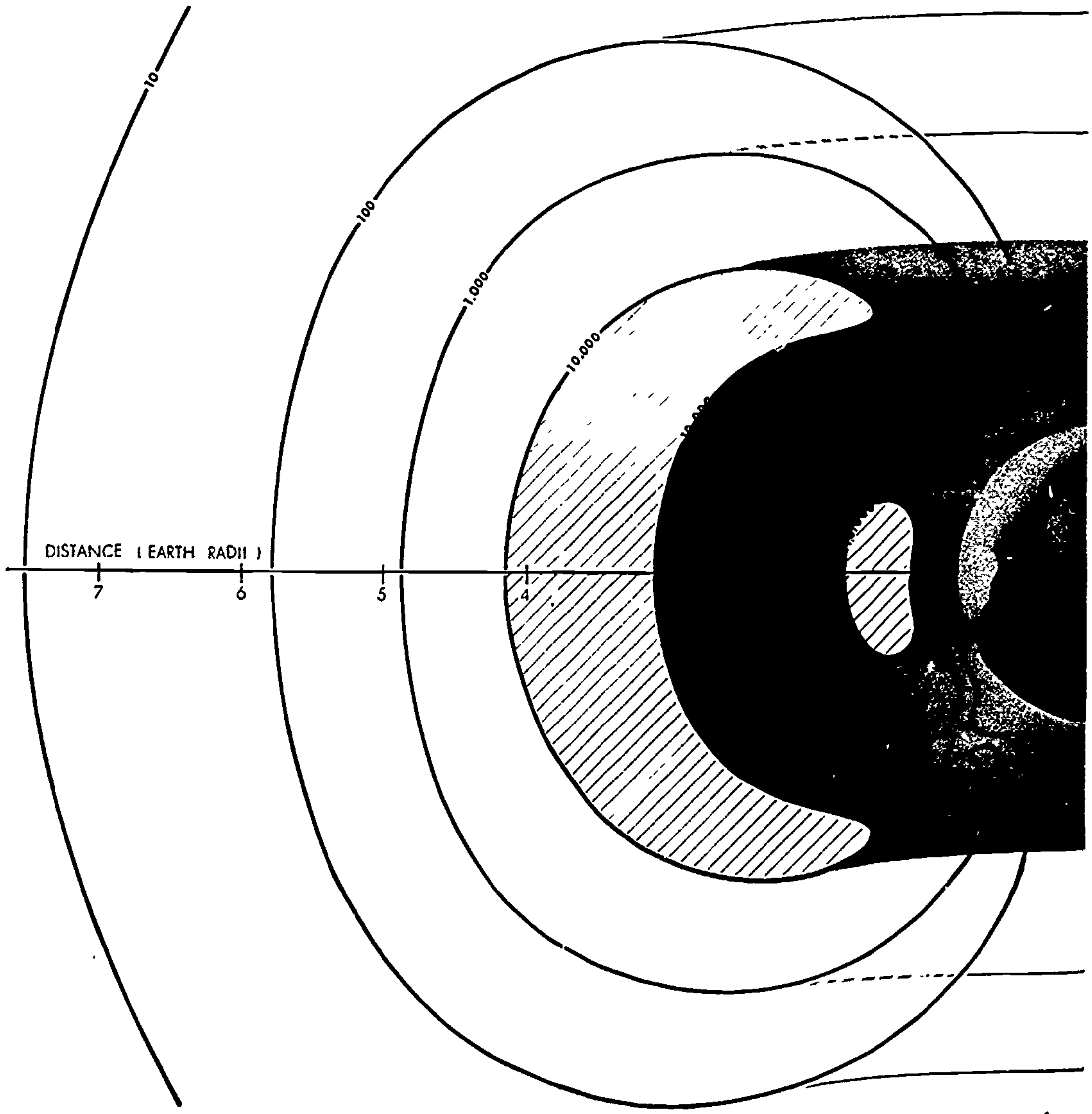
The story of the investigation goes back to 1952 and 1953, before any of us could think realistically about the use of earth satellites to explore the environment of the earth. Parties from our laboratory at the State University of Iowa spent the summers of those years aboard Coast Guard and naval vessels, cruising along a 1,500-mile line from the waters of Baffin Bay, near the magnetic pole in the far northwestern corner of Greenland, southward to the North Atlantic off the coast of Newfoundland. Along the way we launched a series of rocket-carrying balloons—"rockoons." (The balloon lifts a small rocket to an altitude of 12 to 15 miles, whence the rocket carries a modest payload of instruments to a height of 60 to 70 miles.) Our objective was to develop a profile of the cosmic-ray intensities at high altitudes and latitudes, and thus to learn the nature of the low-energy cosmic rays which at lower altitudes and latitudes are deflected by the earth's magnetic field or absorbed in the atmosphere.

Most of the readings radioed down from the rockets were in accord with plausible expectations. Two rockoons sent aloft in 1953, however, provided us with a puzzle. Launched near Newfoundland by Melvin Gottlieb and Leslie Meredith, they encountered a zone of radiation beginning at an altitude of 30 miles that was far stronger than we had expected. At first we were uneasy about the proper operation of our instruments. But critical examination of the data convinced us that we had unquestionably encountered something new in the upper atmosphere.

Significantly these measurements were made in the northern auroral zone. In this zone, which forms a ring some 23 degrees south of the north geomagnetic pole, the incidence of visible auroras reaches its maximum. Since rockets fired north and south of the zone had revealed nothing unusual, we speculated that the strong radiation played some part in the aurora. Showers of particles from the sun, it was thought, come plunging into the atmosphere along magnetic lines of force and set off these displays [see "Aurora and Airglow," by C. T. Elvey and Franklin E. Roach; SCIENTIFIC AMERICAN, September, 1955]. But the theory underlying this explanation did not explain satisfactorily why the aurora and the high-intensity radiation we had detected should occur in the auroral zone and not in the vicinity of the geomagnetic pole itself. Nor could it account for the high energies required to carry the solar particles through the atmosphere to such relatively low altitudes.

The mystery deepened when we found in later studies that the radiation persists almost continuously in the zone above 30 miles, irrespective of visible auroral displays and other known high-altitude disturbances. More discriminating detectors established that the radiation contains large numbers of electrons. Our original observations had detected X-rays only; now it turned out that the X-rays had been generated by the impact of electrons on the skin of the instrument package (as if it had been the "target" in an X-ray tube) and on the sparse atoms of the upper atmosphere itself. Sydney Chapman and Gordon Little at the University of Alaska suggested that such a process might well account for the attenuation of radio signals in the lower ionosphere of the auroral zones.

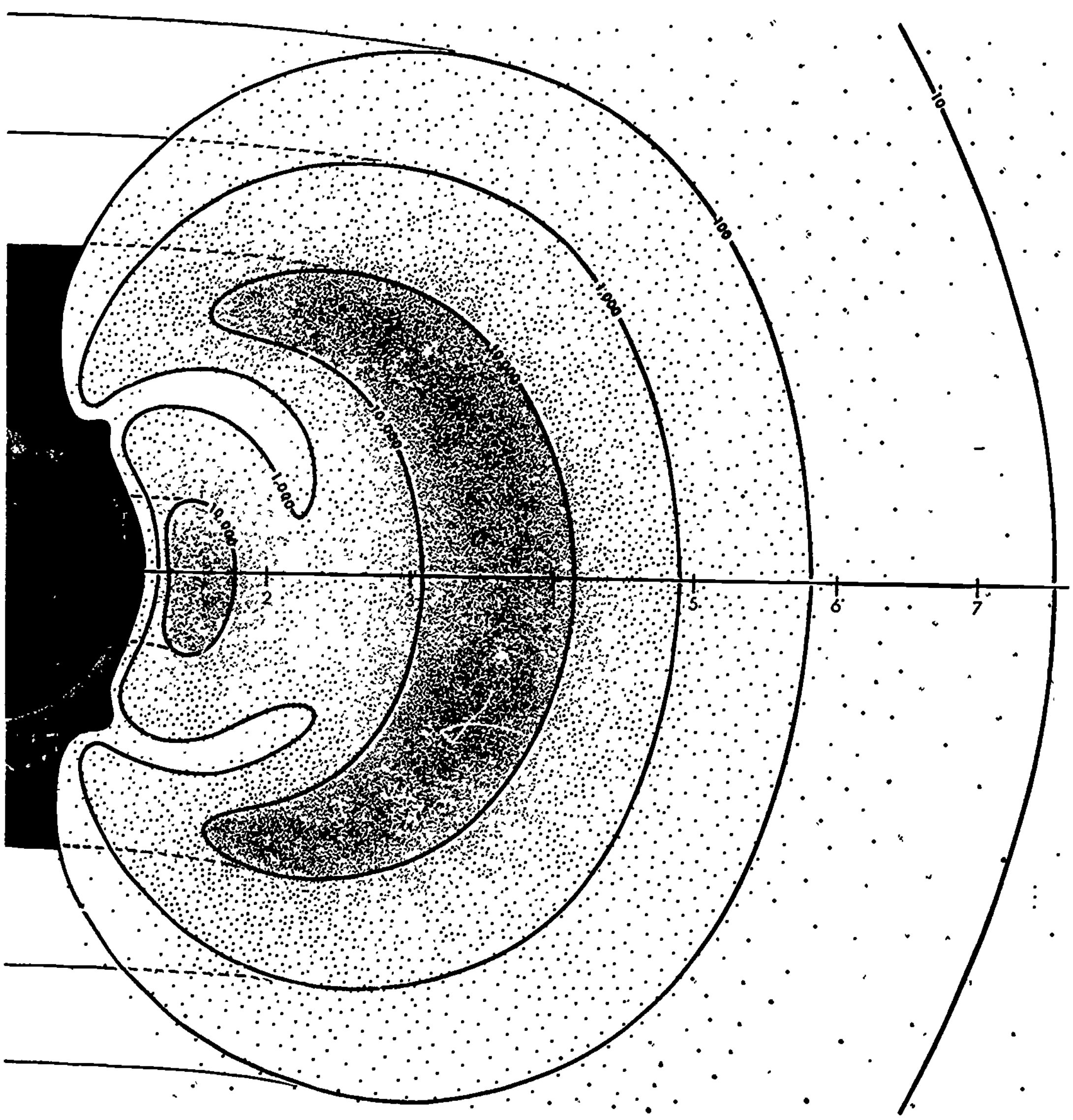
The International Geophysical Year gave us our first opportunity to investigate the "au[illegible]l soft radiation" on a more comprehensive scale. During the



STRUCTURE OF RADIATION BELTS revealed by contours of radiation intensity (*black lines*) is shown schematically by shading (*left*); dots (*right*) suggest distribution of particles in the two belts. Contour numbers give counts per second; horizontal scale

summer and fall of 1957 Laurence Cahill and I launched a number of rockoons off the coast of Greenland and also got off one successful flight in Antarctica. The latter flight established that the radiation exists in the southern as well as the northern auroral zone. In February, 1958, Carl McIlwain fired a series of two-stage rockets through visible auroras above Fort Churchill in Canada, and discovered that the radiation includes energetic protons (hydrogen nuclei) as well as electrons.

Meanwhile all of us had been pushing a new development that greatly expanded the possibilities for high-altitude research. During the summer of 1955 the President and other Government authorities were finally persuaded that it might be feasible to place artificial satellites in orbit, and authorized an I. G. Y. project for this purpose. In January, 1956, a long-standing group of high-altitude experimentalists, called the Rocket and Satellite Research Panel, held a symposium to consider how the satellites could be most fruitfully employed. At that meeting our group proposed two projects. One was to put a satellite into an orbit nearly pole-to-pole to survey the auroral radiation in both the north and south auroral zones. Such orbits, however, did not appear to be

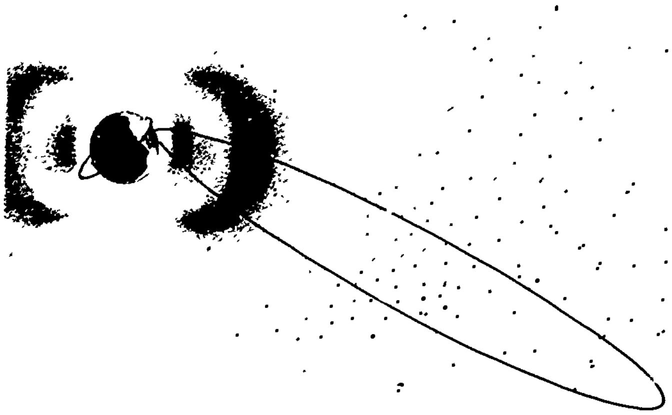


shows distance in earth radii (about 4,000 miles) from the center of the earth. Particles in the inner belt may originate with the radioactive decay of neutrons liberated in the upper atmosphere by cosmic rays; those in the outer belt probably originate in the sun.

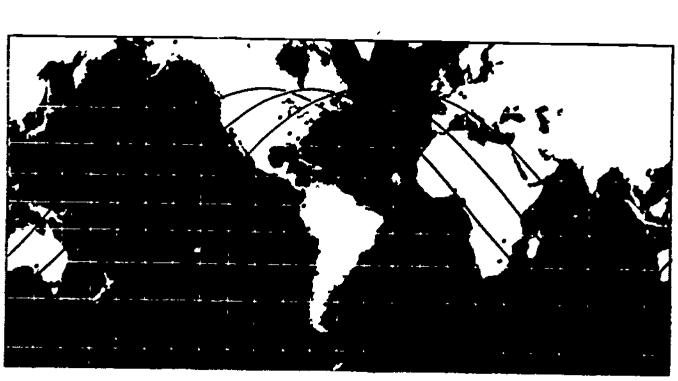
technically feasible in the immediate future. For the time being we were forced to abandon the use of a satellite to probe farther into the auroral soft radiation. We also suggested that a satellite orbiting over the lower latitudes of the earth might usefully be employed in a comprehensive survey of cosmic-ray intensities over those regions. This project was adopted, and we were authorized to prepare suitable experimental apparatus [see "The Artificial Satellite as a Research Instrument," by James A. Van Allen; SCIENTIFIC AMERICAN, November, 1956]. It was planned to place this apparatus on one of the early Vanguard vehicles.

The difficulties and failures of the Vanguard are now history. Sputnik I stimulated some high government officials to accept a proposal that a number of us had been urging for more than a year: to use the proven Jupiter C rocket as a satellite-launching vehicle. As a result on January 31, 1958, Explorer I went into orbit carrying our simple cosmic-ray detector and a radio to broadcast its readings.

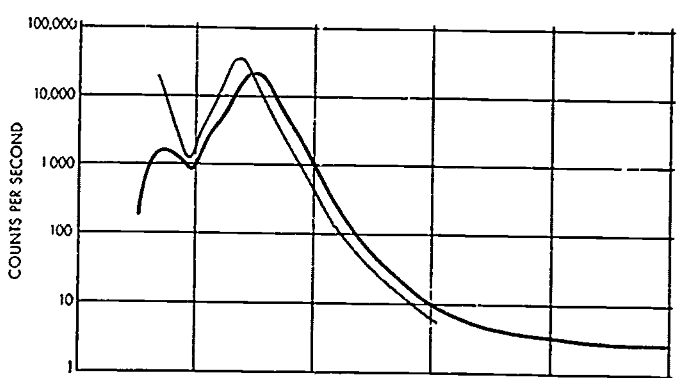
In the first reports from stations located in the U. S. the intensity of radiation increased with altitude along the expected curve. Several weeks later, however, we began to get tapes from stations in



EXPLORER IV AND PIONEER III gave the first detailed picture of the radiation belts. The Explorer IV satellite (*short ellipse*) monitored radiation levels for nearly two months at altitudes up to 1,300 miles. The Pioneer III lunar probe (*long ellipse*) provided data out to 65,000 miles. Its orbit is shown distorted because of the earth's rotation during flight.



EXPLORER IV ORBIT covered the entire region 51 degrees north and south of the equator; the black curve shows a small part of its trace on the earth's surface. More than 25 observation stations (*colored dots*) recorded data from several thousand of the satellite's passes.



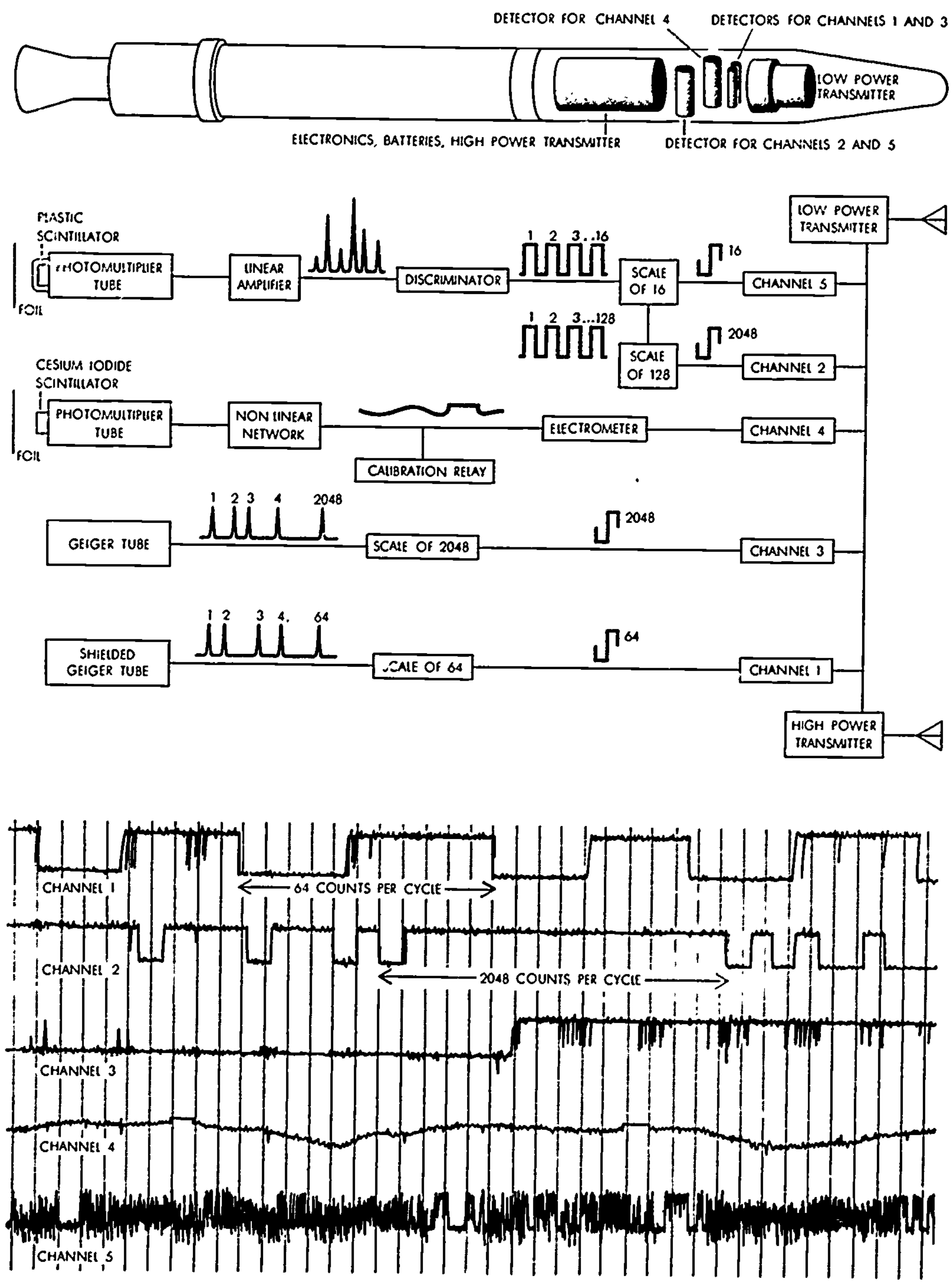
PIONEER III DATA gave the first confirmation of two distinct rings of particles. Counting rates on both the outbound (*black curve*) and the inbound (*gray curve*) legs of the flight showed two peaks. The two curves differ because they cover different sections of the belts.

South America and South Africa which gave us counting rates for much higher altitudes, due to the eccentricity of the satellite's orbit. These records brought us a new surprise. At high altitudes over the equatorial region the apparent counting rate was very low; in some passes it dropped to zero for several minutes. Yet at lower altitudes the rate had quite "reasonable" values—from 30 to 50 counts a second. Again we were uneasy about the trustworthiness of the instruments. The only alternative seemed to be that cosmic rays do not strike the uppermost layers of the atmosphere over the tropics, and we were quite unable to accept this conclusion.

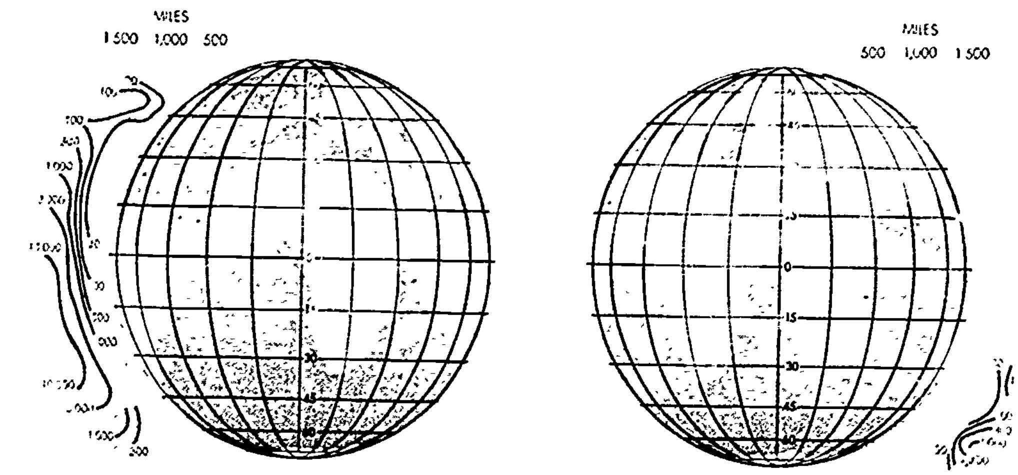
Our uneasiness was increased by the incompleteness of our early data. The Explorer I apparatus broadcast its observations continuously, but its signals could be picked up only intermittently, when the satellite came within range of a ground station. Our original apparatus, designed and developed by George Ludwig for the Vanguard satellites, included a magnetic-tape recorder which could store its observations for a complete orbit around the earth and then report them in a "burst" on radio command from the ground.

By early February, working with the Jet Propulsion Laboratory, we had converted this apparatus for use in the Explorer II satellite. The first attempt to get it into orbit failed. A second rocket placed Explorer III, carrying identical apparatus, in orbit on March 26. This satellite fully confirmed the anomalous results of Explorer I. At altitudes of 200 to 300 miles the counting rate was low. When the satellite went out to 500 to 600 miles, the apparent rate ascended rapidly and then dropped almost to zero. One day, as we were puzzling over the first tapes from Explorer III, McIlwain suggested the first plausible explanation for their peculiar readings. He had just been calibrating his rocket instruments, and called our attention to something that we all knew but had temporarily forgotten: A sufficiently high level of radiation can jam the counter and send the apparent counting rate to zero. We had discovered an enormously high level of radiation, not a lack of it. As Ernest Ray, a member of our group, inaccurately but graphically exclaimed: "Space is radioactive!"

During the next two months Explorer III produced a large number of playback records, every one of which showed the same effect. At low altitudes the counting rate was reasonably attributable to



EXPLORER IV INSTRUMENTS were designed to give a detailed picture of the nature and intensity of the radiation. Plastic scintillator counted only charged particles above certain energies; two different scaling factors adapted it to both high and low counting rates. Cesium-iodide scintillator measured the total energy input rather than individual particles. Shielded and unshielded Geiger tubes could be compared to estimate the penetrability of the radiation. Radio signals suggested by the red curves in upper drawing were recorded by ground stations and later played through a multichannel oscillograph to yield records like that shown below.



TWO SETS OF CONTOURS from readings on opposite sides of the earth *(left and center)* show the northern and southern "horns" of radiation, which point toward the auroral zones: the contour numbers show radiation intensity in counts per second. The "tipped"

cosmic rays. At higher altitudes—the precise height depended on both latitude and longitude—the count increased to very high values. Up to the points at which the counter jammed, it showed counting rates more than 1,000 times the theoretical expectation for cosmic rays. From the rate of increase and the length of the periods of jamming we judged that the maximum count probably went to several times this level. Since the radiation appeared to resemble the auroral soft radiation, we would not have been surprised to find it in the auroral zone or along the magnetic lines of force that connect these zones. But in the equatorial latitudes these lines of force lie much farther out in space than the altitudes attained by the satellites.

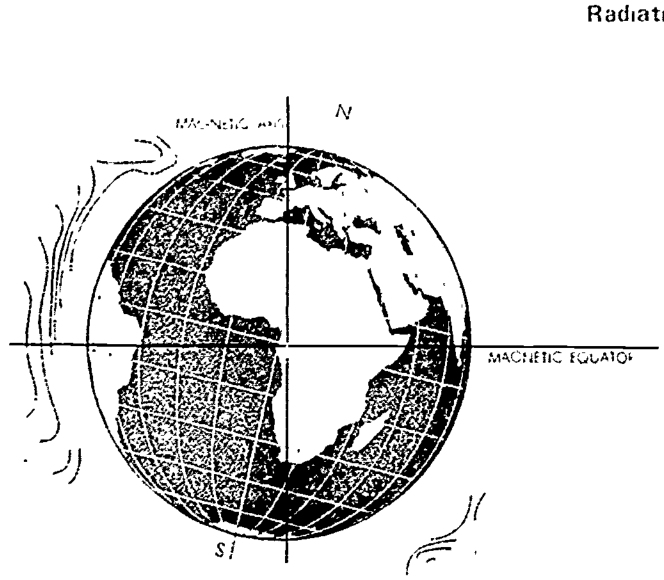
On May 1 of last year we were able to report with confidence to the National Academy of Sciences and the American Physical Society that Explorers I and III had discovered a major new phenomenon: a very great intensity of radiation above altitudes of some 500 miles over the entire region of their traverse, some 34 degrees north and south of the equator. At the same time we advanced the idea that the radiation consists of charged particles—presumably protons and electrons—trapped in the magnetic field of the earth.

We could rule out uncharged particles and gamma and X-rays because they would not be confined by the magnetic field, and so would be observed at lower altitudes. The possibility that the earth's magnetic field might act as a trap for charged particles was first suggested by the Norwegian physicist Carl Stormer in a classical series of papers beginning some 50 years ago, and there was a considerable body of evidence for the existence of low-energy charged particles throughout our solar system and specifically in the vicinity of the earth. But there had been no indication that these particles would possess the high energies we had detected.

From Störmer's theoretical discussion and our own observations we evolved a rough picture of the trapping mechanism. When a fast-moving charged particle is injected into the earth's magnetic field, it describes a corkscrew-shaped trajectory, the center line of which lies along a magnetic line of force. The turns of the helical path are quite open over the equator but become tighter as the particle reaches the stronger magnetic field toward the poles [*see illustration at bottom of opposite page*]. At the lower end of its trajectory the particle goes into a flat spiral and then winds back along a similar path to the other hemisphere, making the transit from one hemisphere to the other in a second or so. During this time its line of travel shifts slightly, so that the particle drifts slowly around the earth as it corkscrews from hemisphere to hemisphere. An electron drifts from west to east; a proton, in the opposite direction. At each end of its path the particle descends into regions of higher atmospheric density; collisions with the atoms of atmospheric gases cause it gradually to change its trajectory and to lose energy. After a period of days or weeks the particle is lost into the lower atmosphere.

There was obviously an urgent scientific need to extend these observations with equipment of greater dynamic range and discrimination. In April of 1958 we persuaded several Federal agencies to support further satellite flights of our radiation equipment as an adjunct to the I. G. Y. program, and we received the enthusiastic support of the National Academy of Sciences for the continuation of our work. We also persuaded the Army Ballistic Missile Agency and the Cape Canaveral Air Force Base to try to place the satellite in an orbit more steeply inclined to the equator; at an inclination of about 50 degrees to the equator it would cover a much greater area of earth and skim the edges of both auroral zones.

Working night and day, we set out at once to build new apparatus of a more discriminating nature. We retained the Geiger tube, which we had used in previous satellites, as a basic "simple-minded" detector. To be ready for the highest intensities of radiation, however, we used a much smaller tube that would yield a lower count in a given flux of radiation, and we hooked it into a circuit that would scale down its count by a much larger factor. To obtain a better idea of the penetrability of the radiation



drawing at right shows the essential symmetry of the radiation around the earth's magnetic axis. The structure of the radiation zone was built up from hundreds of observed points.

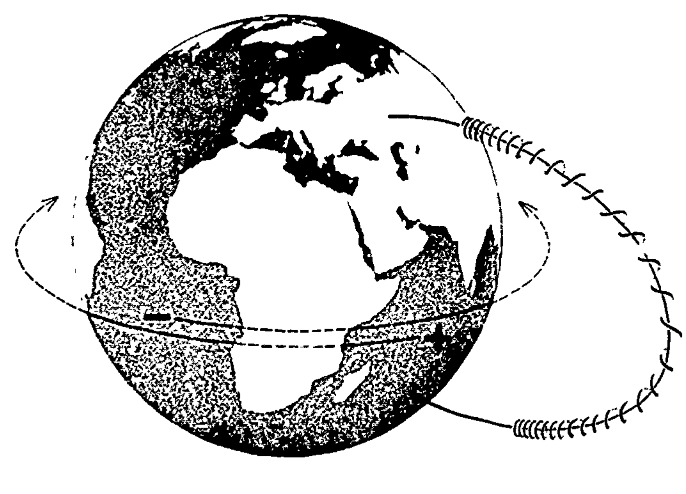
we shielded a similar Geiger tube with a millimeter of lead. As a more discriminating particle detector we adopted a plastic scintillator and photomultiplier tube to respond to electrons with an energy of more than 650,000 electron volts and to protons of more than 10 million electron volts. Finally we glued a thin cesium-iodide crystal to the window of another photomultiplier tube; the light emitted by the crystal when it was irradiated would measure the over-all input of energy rather than the arrival of individual particles. To keep out light when the crystal faced the sun, we shielded it with thin, opaque nickel foil. A special amplifier gave this detector a large dynamic range extending from about .1 erg per second to 100,000 ergs per second.

Explorer IV carried this apparatus into orbit on July 26, and sent down data for almost two months. Magnetic tapes from some 25 observing stations flowed in steadily from late July to late September; altogether we obtained some 3,600 recorded passes of the satellite. A typical pass was readable for several minutes; some of the best were readable for up to 20 minutes, a large fraction of the time required for the satellite to make a turn around the earth. We are still analyzing this mass of data, but the preliminary results have already proved to be enlightening.

The readings have confirmed our earlier estimates of the maximum levels of radiation. Moreover, we have extended our observations to more than 50 degrees north and south of the equator and have been able to plot the intensity of the radiation at various latitudes and longitudes for altitudes up to 1,300 miles. The intensity contours follow the shape of the earth in the equatorial region, but as they approach high northern and southern latitudes they swing outward, then inward and sharply outward again to form "horns" reaching down toward the earth near the auroral zones [*see illustrations at the top of these two pages*]. The entire picture so far is completely consistent with the magnetic-trapping theory.

It was clear from the contours that Explorers I, III and IV penetrated only the lower portion of the radiation belt. As early as last spring we began to make hypothetical extensions of the observed contours out to a distance of several thousand miles. One of these speculative diagrams showed a single, doughnut-shaped belt of radiation with a ridge around the northern and southern edges of its inner circumference, corresponding to the horns of the contours. Another showed two belts—an outer region with a banana-shaped cross section that extended from the northern to the southern auroral zone and an inner belt over the equator with a bean-shaped cross section [*see illustration on pages 40 and 41*]. The latter diagram seemed to fit the contours better. In our seminars and after-hour discussions McIlwain held out for the two-belt theory. The rest of us tended to agree with him but preferred to stay with the single "doughnut" because of its simplicity.

To take the question out of the realm of speculation we had to secure measurements through the entire region of radiation. In May, therefore, I arranged to have one of our radiation detectors carried aboard the lunar probes planned for the fall of 1958. On October



TRAPPED PARTICLES spiral rapidly back and forth along a corkscrew-shaped path whose center is a magnetic line of force. At the same time they drift slowly around the earth (*broken arrows*). Electrons (*negative*) and protons (*positive*) drift in opposite directions.

11, 12 and 13 Pioneer I, the first lunar probe, carried our instruments nearly 70,000 miles out from the earth. Though its readings were spotty, they confirmed our belief that the radiation extended outward for many thousands of miles, with its maximum intensity no more than 10,000 miles above the earth.

The next attempted moon shot, Pioneer II, was a fizzle. Pioneer III, however, went off beautifully on December 6. Although this rocket was intended to reach the vicinity of the moon, we were almost as pleased when it failed to do so, for it gave us excellent data on both the upward and downward legs of its flight, cutting through the radiation region for 65,000 miles in two places.

The observations on both legs showed a double peak in intensity [*see illustration at bottom of page 42*], establishing that there are indeed two belts rather than one. The inner belt reaches its peak at about 2,000 miles from the earth, the outer one at about 10,000 miles. Beyond 10,000 miles the radiation intensity diminishes steadily; it disappears almost completely beyond 40,000 miles. The maximum intensity of radiation in each belt is about 25,000 counts per second, equivalent to some 40,000 particles per square centimeter per second.

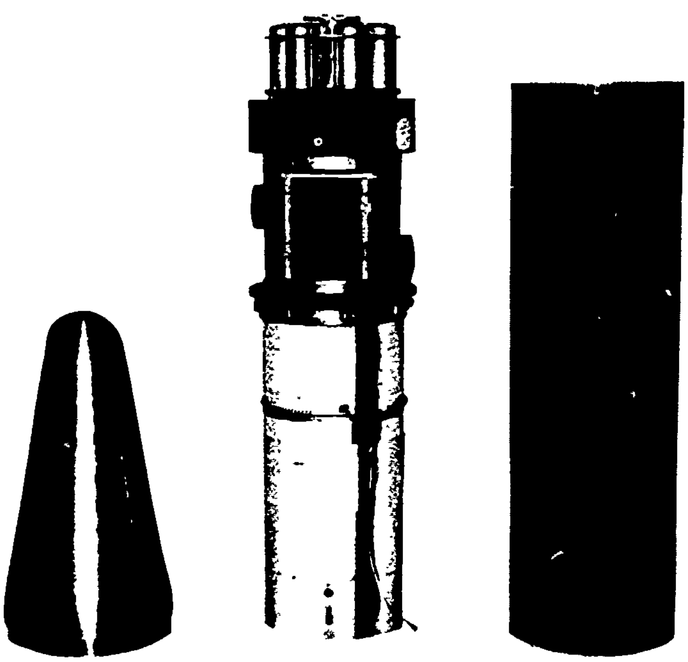
Most of us believe that this great reservoir of particles originates largely in the sun. The particles are somehow injected into the earth's magnetic field, where they are deflected into corkscrew trajectories around lines of force and trapped. In this theoretical scheme the radiation belts resemble a sort of leaky bucket, constantly refilled from the sun and draining away into the atmosphere. A particularly large influx of solar particles causes the bucket to "slop over," mainly in the auroral zone, generating visible auroras, magnetic storms and related disturbances. The normal leakage may be responsible for the airglow which faintly illuminates the night sky and may also account for some of the unexplained high temperatures which have been observed in the upper atmosphere.

This solar-origin theory, while attractive, presents two problems, neither of which is yet solved. In the first place the energy of many of the particles we have observed is far greater than the presumed energy of solar corpuscles. The kinetic energy of solar corpuscles has not been measured directly, but the time-lag between a solar outburst and the consequent magnetic disturbances on earth indicates that the particles are slow-moving and thus of relatively low energy. It may be that the earth's magnetic field traps only a high-energy fraction of the particles. Alternatively, some unknown magnetohydrodynamic effect of the earth's field may accelerate the sluggish particles to higher velocities. Some such process in our galaxy has been suggested as responsible for the great energies of cosmic rays. The second problem in the solar-origin theory is that it is difficult to explain how charged particles can get into the earth's magnetic field in the first place. We believe that neither problem is unsolvable.

Nicholas Christofilos of the University of California and the Soviet physicist S. N. Vernov have suggested an entirely different theory of how the radiation originates. They note that neutrons are released in large numbers in the earth's upper atmosphere by the impact of cosmic rays. These neutrons, being uncharged, can travel through the magnetic field without deflection. In due course some of them decay there into electrons and protons, which are trapped.

Our group agrees that particle-injection of this sort is going on, and at a rate which can be easily calculated; but we feel for a number of reasons that it cannot be the main source of radiation-belt particles. If we are right in supposing that the radiation belts provide the "reservoir" for the aurora, the neutron hypothesis cannot account for more than one 10,000th of the auroral energy output. Even if the association between the radiation belts and the aurora turns out to be fortuitous, preliminary indications both from our work and from the Russian experience with Sputnik III suggest that most of the particles in the radiation belt have much lower energies than those of particles that would be produced by neutron decay. A full knowledge of the energy distribution of the particles will aid greatly in clarifying their origin.

Neither theory explains why there should be two belts rather than one. It is tempting to combine the two theories and suppose that the inner belt originates with "internal injection"—*i.e.*, neutron-decay products—and the outer one with "external injection" of solar corpuscles. The two-belt configuration may of course be a transitory phenomenon, though the data from Explorer IV and Pioneer III indicate that the separate belts persisted in essentially the same form for at least five months. We should bear in mind, however, that 1958 was a year of great solar activity. Three years



HEAD OF EXPLORER IV includes nose cone (*left*), instrument "payload" (*center*) and protective shell (*right*). Payload includes four detectors, two radio transmitters, batteries and associated electronic circuity. The outer shell is approximately six inches in diameter.

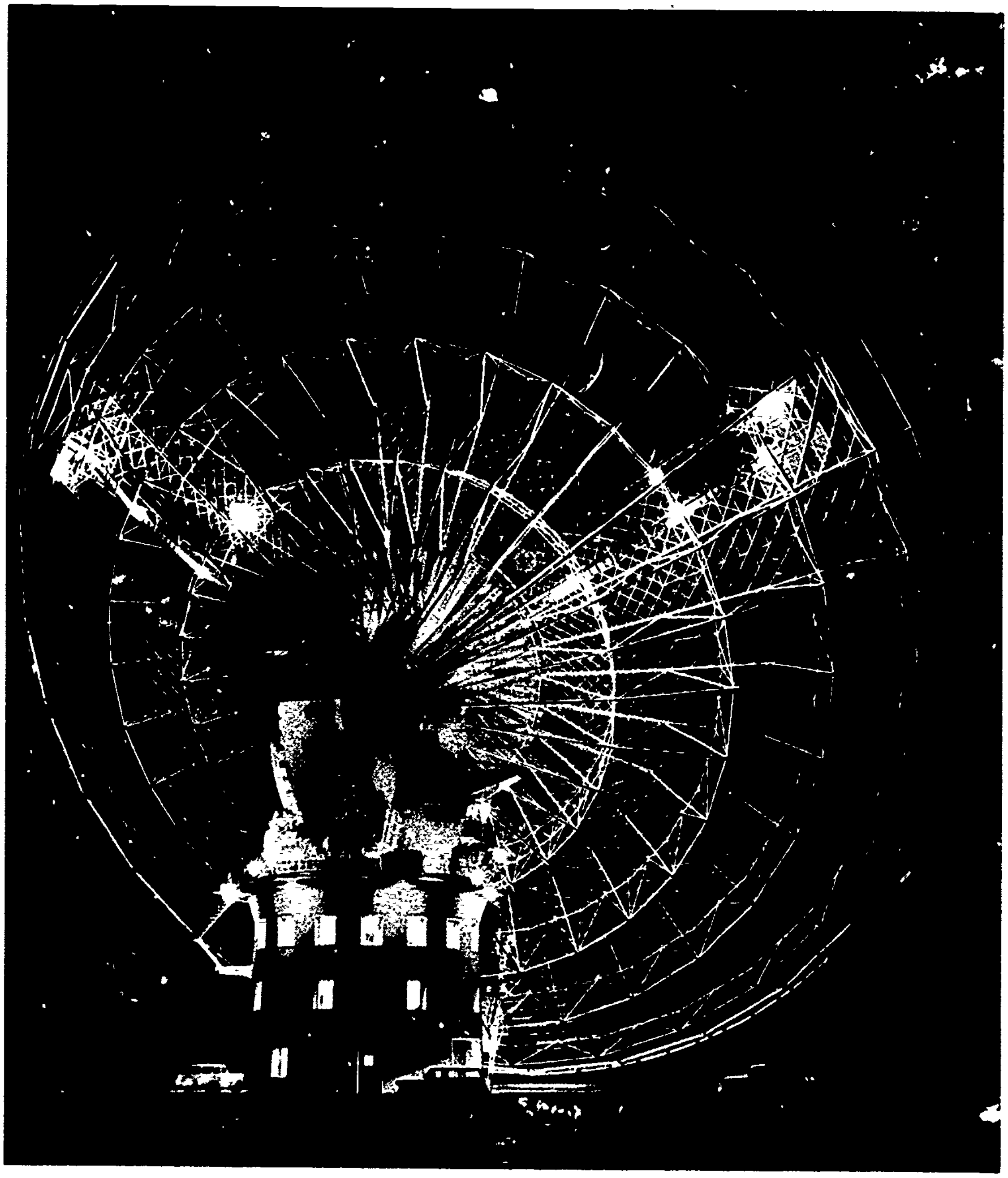
FOUR-STAGE ROCKET launched the Pioneer III moon probe on December 6, 1958. Though the flight failed to reach the moon, its outbound leg gave a continuous record of radiation out to 65,000 miles; the inbound leg gave data between 30,000 and 10,000 miles.

from now we may well find a much lower over-all intensity and perhaps a different structure altogether.

In addition to these possible long-term changes, there may be short-term fluctuations in the belts. While we feel sure that the influx and leakage of particles must balance in the long run, a major solar outbreak may temporarily increase the intensity of the radiation many-fold. If we were to detect such fluctuations and were to find that they coincide with solar outbursts on the one hand and with terrestrial magnetic disturbances on the other, we would have a plain lead to the origin of the particles. Before long we hope to launch a satellite that will monitor radiation levels for at least a year.

Our measurements show that the maximum radiation level as of 1958 is equivalent to between 10 and 100 roentgens per hour, depending on the still-undetermined proportion of protons to electrons. Since a human being exposed for two days to even 10 roentgens would have only an even chance of survival, the radiation belts obviously present an obstacle to space flight. Unless some practical way can be found to shield space-travelers against the effects of the radiation, manned space rockets can best take off through the radiation-free zone over the poles. A "space station" must orbit below 400 miles or beyond 30,000 miles from the earth. We are now planning a satellite flight that will test the efficacy of various methods of shielding.

The hazard to space-travelers may not end even when they have passed the terrestrial radiation belts. According to present knowledge the other planets of our solar system may have magnetic fields comparable to the earth's and thus may possess radiation belts of their own. The moon, however, probably has no belt, because its magnetic field appears to be feeble. Lunar probes should give us more definite information on this point before long.



Parkes radio telescope in the Goobang Valley, New South Wales, Australia.

How does the brain work? Part of the answer lies in electrophysiology, the study of the relation between electricity and nervous stimulation.

17 A Mirror for the Brain

W. Grey Walter

A chapter from his book *The Living Brain* published in 1963.

The Greeks had no word for it. To them the brain was merely "the thing in the head," and completely negligible. Concerned as so many of them were about man's possession of a mind, a soul, a spiritual endowment of the gods, it is strange they did not anticipate our much less enterprising philosophers of some score of centuries later, and invent at least a pocket in the head, a sensorium, to contain it. But no, the Greeks, seeking a habitation for the mind, could find no better place for it than the midriff, whose rhythmic movements seemed so closely linked with what went on in the mind

The Hebrews also attributed special dignity to that part of the body; thence Jehovah plucked man's other self. Old ideas are not always as wide of the mark as they seem. The rhythm of breathing is closely related to mental states. The Greek word for diaphragm, *phren*, appears in such everyday words as *frenzy* and *fran*tic, as well as in the discredited *phreno*logy and the erudite schizo*phrenia*.

Above the midriff the classical philosophers found the vapours of the mind; below it, the humours of the feelings. Some of these ideas persisted in physiological thought until the last century and survive in the common speech of today. Hysteric refers by derivation to the womb. The four basic human temperaments were: choleric, referring to the gall bladder; phlegmatic, related to inflammation; melancholic, black bile; and sanguine, from the blood. This classification of temperaments was revived by a modern physiologist, Pavlov, to systematize his observations of learning.

As in nearly all notions that survive as long as these fossils of language have survived, there is an element of truth, of observation, in them. States of mind are certainly related to the organs and liquors designated, and may even be said in a sense to originate in them. The philosopher, William James, was responsible with Lange for a complete theory of emotion which invoked activity in the viscera as the essential precursor of deep feeling. Some of the most primitive and finest phrases in English imply this dependence of sincere or deep emotion on heart or bowels. But communication of thought is so rapid that the Greeks overlooked the existence or need of a relay station. And no doubt it is for the same reason that we all seem particularly given to the same error of over-simplification when we first begin, or refuse to begin, to consider how the mind works. We know what makes us happy or unhappy. Who, in the throes of sea-sickness, would think of dragging in the brain to account for his melancholy state?

More curious still is Greek negligence of the brain, considering their famous oracular behest, "Know thyself." Here indeed was speculation, the demand for a mirror, insistence upon a mirror. But for whom, for what? Was there, among the mysteries behind the altar, concealed perhaps in the Minerva myth, a suspicion of something more in the head than a thing, and that the organ which had to do the knowing of itself must be an organ of reflection?

The brain remained for more than two thousand years in the dark after its coming of age. When it was discovered by the anatomist, he explored it as a substance in which might be found the secret dwelling of intelligence; for by that time the mind had moved from the diaphragm to the upper story, and Shakespeare had written of the brain, "which some suppose the soul's frail dwelling-house." Dissection was high adventure in those days. Most people believed what an ironical writer today was "astonished to learn," that "it is possible for anger, envy, hatred, malice, jealousy, fear and pride, to be confined in the same highly perishable form of matter with life, intelligence, honesty, charity, patience and truth." The search for such prize packets of evil and virtue in the brain tissue, dead or alive, could only lead to disappointment. The anatomist had to be satisfied with weighing the "grey matter"—about 50 ounces for man and 5 less for woman—and making sketches of the very complicated and indeed perishable organisation of nerves and cells which his knife revealed. He could do little more. It should enlighten us at once as to the essential character of brain activity, that there was no possible understanding of the mechanism of the brain until the key to it, the electrical key, was in our hands.

There were some flashes of foresight, sparks in the scientific dark, before Galvani put his hand on the key. What generated all the speculations of the day was a new notion in science, the conception of physical motion which began to acquire importance with Galileo and continued with Newton and into our own times with Rutherford and Einstein. First among these imaginative flashes may be mentioned the novel proposal made by the 16th Century philosopher, Hobbes, when disputing the dualist theory of Descartes. The French philosopher contemplated a non-spatial mind influencing the body through the brain, and suggested the pineal gland as the rendezvous for mind and matter. The proposal advanced by Hobbes, in rejecting this popular theory, was that thought should be regarded as being produced by bodies in motion. Hobbes was born in the year of the Spanish Armada; the Royal Society had received its charter seventeen years before he died in 1679.

The controversy about the residential status of the mind is almost as much out of date as that in which the non-existence of motion seemed to be proved by the hare and tortoise fable. But the value of Hobbes' speculation was enduring; the observation and correlation of mental and physical phenomena are today a routine of physiological research.

More specific than the speculation of Hobbes was that of Dr. David Hartley about a century later. Hartley in 1749 anticipated by two hundred years the kind of theory of mental function for which evidence has been found in the last year or two. His "Observations on Man, his Frame, his Duty and his Expectations" is a milestone in the history of English thought. Hartley, a contemporary of Newton and Hume, was a pioneer of what he termed the "doctrine of mechanism." According to this, he suggested, mental phenomena are derived from rhythmic movements in the brain—vibrations, he called them; upon these is superimposed a fine structure of "vibratiuncles" which give thought and personality their subtle shades and variations. Hartley realised quite well the value of the plastic and compact virtues such a system might have. He was also the first to develop the theory of "association of ideas" in a rigorous form, relating this to his "vibratiuncles" in a manner which we should now consider strictly scientific in the sense that it is susceptible to experimental test. It is difficult for us to appreciate the originality of his notions, the gist of which is now a commonplace of electrophysiology.

Hartley wrote nearly half a century before Galvani (1737–1798) and with him we might say farewell to fancy. But to pass over the famous Galvani-Volta controversy with the bald statement that the one claimed to have discovered electricity in animals and the other its generation by metals, would be unfair to any reader who may not know how strangely truth came out of that maze of error.

The incident began with an experiment made by Luigi and Lucia Galvani in the course of their long and patient study of that still fresh mystery, electricity. The word had been in use since William Gilbert coined it in the 16th century from *elektron*, meaning amber, another pretty semantic shift; and Henry Cavendish had already, eight years before the incident, determined the identity of its dynamic laws with those of gravitation. Everybody in high society was familiar with the effects of discharges from Leyden jars upon the lifeless muscles of executed criminals; and Louis XV had, in the words of Silvanus Thomson, "caused an electric shock from a battery of Leyden jars to be administered to 700 Carthusian monks joined hand to hand, with prodigious effect." But in Bologna in 1790 the professor of anatomy had a notion that it was atmospheric electricity which acted upon the muscle tissues of animals. On a stormy evening, one version of the story goes, he and his wife had the curious idea of testing this point by tying a dead frog to the top of the iron balustrade of the court-yard, apparently using copper wire to hold it by the leg. They expected that, as the storm approached, the frog would be convulsed by electric shocks. And, as they watched the thunder cloud come near, so indeed it happened; the dead frog, hanging against the iron grill, twitched in repeated convulsions.

Further experiments convinced the Galvani that they had witnessed a form of electricity derived from living processes, not merely from the atmosphere. He published a famous account of his experiments on the relation of animal tissue to electricity: *De viribus Electricitatis in Motu Musculari Commentarius* (1791). Volta seized upon this to refute the whole of Galvani's thesis, repeating his experiment not only without the storm but without the frog, proving that the electricity in question could be generated by copper and zinc sheets. This "current electricity" as it was called, was therefore metallic, and no nonsense about any animal variety. So ended a controversy and a friendship. So began the science of electrical engineering.

Eppur, the Galvani might have repeated, *si muove*. For their discredited experiment had truly revealed, not indeed what they supposed, but something more wonderful. What had happened was that, swaying in the wind, the suspended frog had come into contact with the iron bars, between which and the copper wire a current had been generated, activating its muscles. The Galvani had demonstrated the electrical aspect of nervous stimulation.

This was an event as important to the physiologist as its counter-event was to the physicist; it was the starting-point of that branch of the science with which we are concerned here, electrophysiology.

Volta's counter-demonstration led directly to the invention of the electric battery, and economic opportunity evoked electrical engineering from the Voltaic pile. There was no such incentive for research when, a generation later, the existence of animal electricity was proved. Instead, the discovery was exploited by the academic dilettante and the quack. The Aristotelian doctors of the period, assuming that where there is electricity there is magnetism, saw in it proof also of Mesmer's *"Propositions"* which had been published in his *"Mémoire sur la Découverte du Magnétisme Animal"* in 1779, floundering deeper into mystification than Dr. Mesmer himself, who had at least declared in his *"Mémoire"* that he used the term analogically, and that he "made no further use of electricity or the magnet from 1776 onwards."

There is still controversy about the origin and nature of animal electricity. Nobody who has handled an electric eel will question the ability of an animal to generate a formidable voltage; and the current is demonstrably similar in effect to that of a mineral dry cell. On the other hand, there is no evidence that the electric energy in nerve cells is generated by electro-magnetic induction or by the accumulation of static charge. The bio-chemist finds a complicated substance, acetylcholine, associated with electric changes; it would be reasonable to anticipate the presence of some such substance having a role at least as important as that of the chemicals in a Leclanché cell.

We know that living tissue has the capacity to concentrate potassium and distinguish it from sodium, and that neural electricity results from the differential permeability of an inter-face, or cell-partition, to these elements, the inside of a cell being negatively charged, the outside positively. Whether we call this a chemical or an electrical phenomenon is rather beside the point. There would be little profit in arguing whether a flash-lamp is an electrical or chemical device; it is more electrical than an oil lamp, more chemical than a lightning flash. We shall frequently refer to changes of potential as electrical rhythms, cycles of polar changes, more explicitly electro-chemical changes. We shall be near the truth if we keep in mind that electrical changes in living tissue, the phenomena of animal electricity, are signs of chemical events, and that there is no way of distinguishing one from the other in the animal cell or in the mineral cell. The current of a nerve impulse is a sort of electro-chemical smoke-ring about two inches long travelling along the nerve at a speed of as much as 300 feet per second.

The neglect and mystification which obscured Galvani's discovery, more sterile than any controversy, forced electrophysiology into an academic backwater for some decades. A few experiments were made, for example, by Biedermann, who published a 2-volume treatise called *Electrophysiology*, and by Dubois-Reymond, who introduced Michael Faraday's induction coil into the physiological laboratory and the term faradisation as an alternative to galvanisation into the physiotherapist's vocabulary. Faraday's electrical and electrifying research began in 1831, the date also of the foundation of the British Association for the Advancement of Science; but physiology long remained a backward child of the family.

Hampered though these experimenters were by lack of trustworthy equipment—they had to construct their own galvanometers from first principles—they gradually accumulated enough facts to show that all living tissue is sensitive in some degree to electric currents and, what is perhaps more important, all living tissue generates small voltages which change dramatically when the tissue is injured or becomes active.

These experiments were not concerned with the brain; they were made on frog's legs, fish eggs, electric eels and flayed vermin. Nor could the brain be explored in that way.

> Following life through creatures you dissect,
> You lose it in the moment you detect.

It took a war to bring the opportunity of devising a technique for exploring the human brain—and two more wars to perfect it. Two medical officers of the Prussian army, wandering through the stricken field of Sedan, had the brilliant if ghoulish notion to test the effect of the Galvanic current on the exposed brains of some of the casualties. These pioneers of 1870, Fritsch and Hitzig, found that when certain areas at the side of the brain were stimulated by the current, movements took place in the opposite side of the body.

That the brain itself produces electric currents was the discovery of an English physician, R. Caton, in 1875.

This growing nucleus of knowledge was elaborated and carried further by Ferrier in experiments with the "Faradic current." Toward the end of the century there was a spate of information which suggested that the brain of animals possessed electrical properties related to those found in nerve and muscle. Prawdwicz-Neminski in 1913 produced what he called the "electro-cerebrogram" of a dog, and was the first to attempt to classify such observations.

The electrical changes in the brain, however, are minute. The experiments of all these workers were made on the exposed brains of animals. There were no means of amplification in those days, whereby the impulses reaching the exterior of the cranium could be observed or recorded, even if their presence had been suspected. On the other hand, the grosser electrical currents generated by the rhythmically contracting muscles of the heart were perceptible without amplification. Electro-cardiography became a routine clinical aid a generation before the invention of the thermionic tube made it possible to study the electrical activity of the intact human brain.

From an unexpected quarter, at the turn of the century, came an entirely new development. Turn up the section on the brain in a pre-war textbook of physiology and you will find gleanings from clinical neuro-anatomy and—Pavlov. Almost as if recapitulating the history of physiological ideas, Pavlov's work began below the midriff. He found that the process of digestion could not be understood without reference to the nervous system, and so commenced his laborious study of learning in animals.

In the gospel according to Stalin, Pavlov founded not merely a branch of physiology as Galvani had done, but a whole new science—Soviet physiology. His work indeed was original; it owed nothing to Galvani, lying quite outside electrophysiology, to which it was nevertheless eventually, though not in Pavlov's day, to contribute so much in the way of understanding.

For nearly two generations Pavlov's experiments were the major source of information on brain physiology. Workers in the English laboratories had not permitted themselves to explore further than the top of the spinal cord. One took an anatomical glance at the brain, and turned away in despair. This was not accountable to any peculiar weakness of physiological tradition but to the exigencies of scientific method itself. A discipline had been building up through the centuries which demanded that in any experiment there should be only one variable and its variations should be measurable against a controlled background. In physiology this meant that in any experiment there should be only one thing at a time under investigation—one single function, say, of an organ—and that the changes of material or function should be measurable. There seemed to be no possibility of isolating one single variable, one single mode of activity, among the myriad functions of the brain. Thus there was something like a taboo against the study of the brain. The success of Pavlov in breaking this taboo early in the century was due to his contrivance for isolating his experimental animals from all but two stimuli; his fame rests on his measurement of responses to the stimuli.

There was no easy way through the academic undergrowth of traditional electrophysiology to the electrical mechanisms underlying brain functions. The Cambridge school of electrophysiology, under a succession of dexterous and original experimenters beginning toward the end of the last century, developed its own techniques in special fields of research, particularly in the electrical signs of activity in muscles, nerves and sense organs. At the same time, the Oxford school under the leadership of Sherrington was beginning to unravel some of the problems of reflex function of the spinal cord. In both these schools the procedure adopted, to comply with the traditional requirements of scientific method, was to dissect out or isolate the organ or part of an organ to be studied. This was often carried to the extreme of isolating a single nerve fibre only a few thousandths of a millimetre in diameter, so as to eliminate all but a single functional unit.

Imagine, then, how refreshing and tantalizing were the reports from Pavlov's laboratory in Leningrad to those engaged on the meticulous dissection of invisible nerve tendrils and the analysis of the impulses which we induced them to transmit. After four years spent working literally in a cage and chained by the ankle—not for punishment but for electrical screening—enlargement came when my professor of that date, the late Sir Joseph Barcroft, assigned me to establishing a laboratory in association with a visiting pupil of Pavlov, Rosenthal. We spent a year or so on mastering the technique and improving it by the introduction of certain electronic devices. The Russian results were confirmed. To do more than this would have required staff and equipment far beyond the resources of the Cambridge laboratory.

Meanwhile, another major event in the history of physiology had taken place. Berger, in 1928, at last brought Hartley's vibrations into the laboratory and with them a method which seemed to hold out the promise of an investigation of electrical brain activity as precise as were the reflex measurements of Pavlov. When Pavlov visited England some time after we heard of this, as the English exponent of his work I had the privilege of discussing it with him on familiar terms. Among other things, I asked him if he saw any relation between the two methods of observing cerebral activity, his method and Berger's. The latter, I was even then beginning to suspect, might in some way provide a clue to *how* the conditioning of a reflex was effected in the brain. But Pavlov showed no desire to look behind the scenes. He was not in the least interested in the mechanism of cerebral events; they just happened, and it was the happening and its consequence that interested him, not how they happened. Soviet physiology embalmed the body of this limited doctrine as mystically as the body of Lenin, for the foundations of their science. The process of conditioning reflexes has a specious affinity with the Marxian syllogism. Others have found in the phenomena sufficient substantiation for a gospel of Behaviourism.

Pavlov was before his time. He would have been a greater man, his work would have been more fertile in his lifetime, and Russian science might have been spared a labyrinthine deviation, had the work of Berger come to acknowledgement and fruition in his day. But again there was delay; Berger waved the fairy wand in 1928; the transformation of Cinderella was a process of years.

There were reasons for this delay. For one thing, Berger was not a physiologist and his reports were vitiated by the vagueness and variety of his claims and the desultory nature of his technique. He was indeed a surprisingly unscientific scientist, as personal acquaintance with him later confirmed.

The first occasion on which the possibilities of clinical electroencephalography were discussed in England was quite an informal one. It was in the old Central Pathological Laboratory at the Maudsley Hospital in London, in 1929. The team there under Professor Golla was in some difficulty about electrical apparatus; they were trying to get some records of the "Berger rhythm," using amplifiers with an old galvanometer that fused every time they switched on the current. Golla was anxious to use the Matthews oscillograph, then the last word in robust accuracy, to measure peripheral and central conduction times. I was still working at Cambridge under the watchful eye of Adrian and Matthews and was pleased to introduce this novelty to him and at the same time, with undergraduate superiority, put him right on a few other points. When, at lunch around the laboratory table, he referred to the recent publication of Berger's claims, I readily declared that anybody could record a wobbly line, it was a string of artefacts, even if there were anything significant in it there was nothing you could measure, and so on. Golla agreed with milder scepticism, but added: "If this new apparatus is as good as you say, it should be easy to find out whether Berger's rhythm is only artefact; and if it isn't, the frequency seems remarkably constant; surely one could measure that quite accurately." And he surmised that there would be variations of the rhythm in disease.

Cambridge still could not accept the brain as a proper study for the physiologist. The wobbly line did not convince us or anybody else at that time. Berger's "elektrenkephalograms" were almost completely disregarded. His entirely original and painstaking work received little recognition until in May, 1934, Adrian and Matthews gave the first convincing demonstration of the "Berger rhythm" to an English audience, a meeting of the Physiological Society at Cambridge.

Meanwhile, Golla was reorganising his laboratory, and his confidence in the possibilities of the Berger method was growing. When he invited me to join his research team as physiologist at the Central Pathological Laboratory, my first task was to visit the German laboratories, including particularly that of Hans Berger.

Berger, in 1935, was not regarded by his associates as in the front rank of German psychiatrists, having rather the reputation of being a crank. He seemed to me to be a modest and dignified person, full of good humour, and as unperturbed by lack of recognition as he was later by the fame it eventually brought him. But he had one fatal weakness: he was completely ignorant of the technical and physical basis of his method. He knew nothing about mechanics or electricity. This handicap made it impossible for him to correct serious shortcomings in his experiments. His method was a simple adaptation of the electrocardiographic technique by which the electrical impulses generated by the heart are recorded. At first he inserted silver wires under the subject's scalp; later he used silver foil bound to the head with a rubber bandage. Nearly always he put one electrode over the forehead and one over the back of the head; leads were taken from these to an Edelmann galvanometer, a light and sensitive "string" type of instrument, and records were taken by an assistant photographer. A potential change of one-ten-thousandth of a volt—a very modest sensitivity by present standards—could just be detected by this apparatus. Each record laboriously produced was equivalent to that of two or three seconds of modern continuous pen recording. The line did show a wobble at about 10 cycles per second. (See Figure 3.) He had lately acquired a tube amplifier to drive his galvanometer, and his pride and pleasure in the sweeping excursions of line obtained by its use were endearing.

Berger carried the matter as far as his technical handicap permitted. He had observed that the larger and more regular

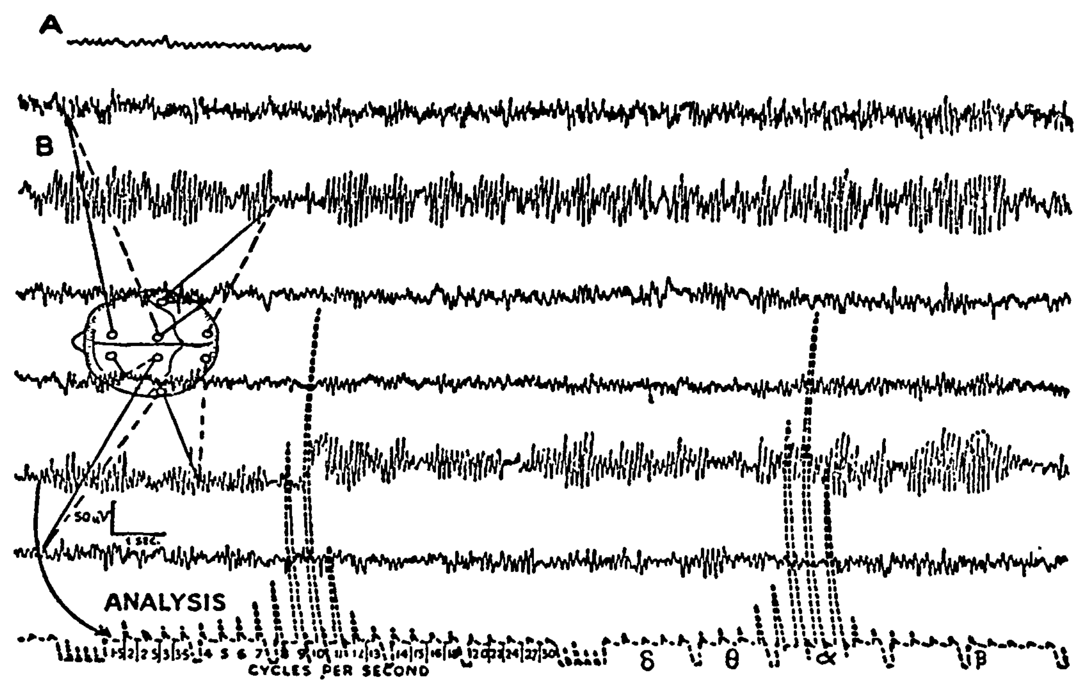


Figure 3. "The line did show a wobble at about 10 cycles per second." (a) A tracing from one of Berger's earliest records. (b) Record from a modern laboratory showing consistency of automatic analysis over 10-second periods.

rhythms tended to stop when the subject opened his eyes or solved some problem in mental arithmetic. This was confirmed by Adrian and Matthews with leads from electrodes on Adrian's hea. tached to a Matthews amplifier and ink-writing oscillograph. This superior apparatus, and a more careful location of electrodes, enabled them to go a step further and prove that the 10 cycles per second rhythm arises in the visual association areas in the occiput and not, as Berger supposed, from the whole brain.

Only some years later was it realised what an important step this was. Its significance could not be recognised while so little was known about the components of the "wobbly line," the electroencephalogram or, abbreviated, EEG. Unavoidably at the time, the significance of the salient character of the normal EEG was overlooked; it was found, in Adrian's phrase, "disappointingly constant." The attention of many early workers in electroencephalography therefore turned from normal research to the study of nervous disease. In immediate rewards this has always been a rich field. In this instance, a surprising state was soon reached wherein what might be called the electropathology of the brain was further advanced than its electrophysiology.

In the pathological laboratory, Golla's earlier surmise, that there would be variations of the rhythmic oscillation in disease, was soon verified. A technique was developed there by which the central point of the disturbance in the tissue could be accurately determined. For surgery, the immediate result of perfecting this technique was important; it made possible the location of tumours, brain injuries, or other physical damage to the brain. It was helpful in many head cases during the war as well as in daily surgical practice.

The study of epilepsy and mental disorders also began to occupy the attention of many EEG workers. The difficulties encountered in these subjects threw into prominent relief the essential complexity of the problem as compared with those of classical physiology. The hope of isolating single functions had now been abandoned; those who entered this field were committed to studying the brain as a whole organ and through it the body as a whole organism. They were therefore forced to multiply their sources of information.

It is now the general EEG practice, not only for clinical purposes, but in research, to use a number of electrodes simultaneously, indeed as many as possible and convenient. The standard make of EEG recorder has eight channels. Eight pens are simultaneously tracing lines in which the recordist, after long experience, can recognise the main components of a complex graph. The graphs can also be automatically analysed into their component frequencies. A more satisfactory method of watching the electrical changes in all the main areas, as in a moving picture, a much more informative convention than the drawing of lines, has been devised at the Burden Neurological Institute. This will be described after a simple explanation of what is meant by the rhythmic composition of the normal EEG; for its nature, rather than the methods of recording and analysing it, is of first importance for understanding what follows.

If you move a pencil amply but regularly up and down on a paper that is being drawn steadily from right to left, the result will be a regular series of curves. If at the same time the paper is moving up and down, another series of curves will be added to the line drawn. If the table is shaking, the vibration will be added to the line as a ripple. There will then be three components integrated in the one wavy line, which will begin to look something like an EEG record. The line gives a coded or conventional record of the various frequencies and amplitudes of different physical movements. In similar coded or integrated fashion the EEG line reports the frequencies and amplitudes of the electrical changes in the different parts of the brain tapped by the electrodes on the scalp, their minute currents being relayed by an amplifier to the oscillograph which activates the pens.

All EEG records contain many more components than this; some may show as many as 20 or 30 at a time in significant sizes. Actually there may be tens of thousands of impulses woven together in such a manner that only the grosser combinations are discernible.

A compound curve is of course more easily put together than taken apart. (See Figure 4.) The adequate analysis of a few inches of EEG records would require the painstaking computation of a mathematician—it might take him a week or so. The modern automatic analyser in use in most laboratories writes out the values of 24 components every 10 seconds, as well as any averaging needed over longer periods.

The electrical changes which give rise to the alternating currents of variable frequency and amplitude thus recorded arise in the cells of the brain itself; there is no question of any other power supply. The brain must be pictured as a vast aggregation of electrical cells, numerous as the stars of the Galaxy, some 10 thousand million of them, through which surge the restless tides of our electrical being relatively thousands of times more potent than the force of gravity. It is when a million or so of these cells repeatedly fire together that the rhythm of their discharge becomes measureable in frequency and amplitude.

What makes these million cells act together—or indeed what causes a single cell to discharge—is not known. We are still a long way from any explanation of these basic mechanics of the brain. Future research may well carry us, as it has car-

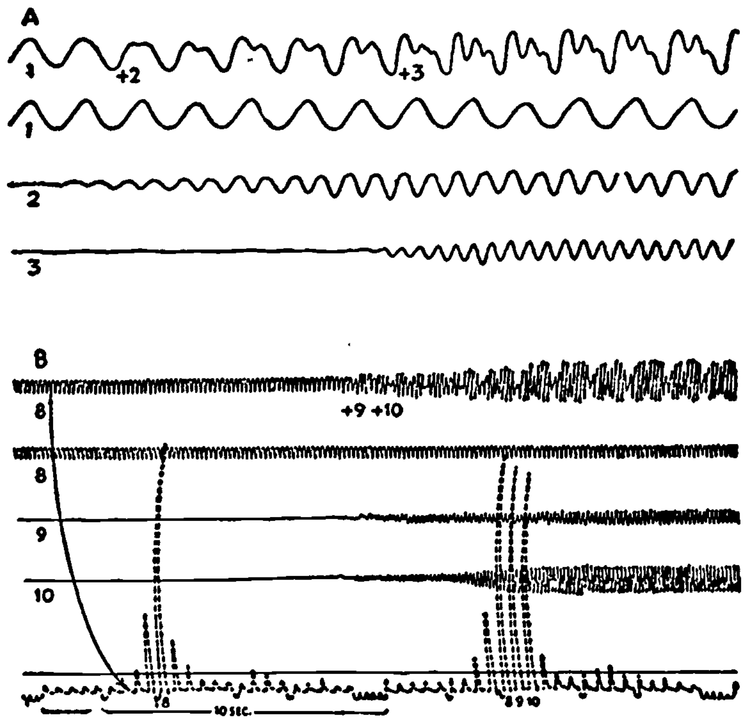


Figure 4. "A compound curve is more easily put together than taken apart." (a) A compound curve in which the three components can be detected by visual inspection, ratios 1:2 and 2:3. (b) The three components (ratios 8:9, 9:10) of this compound curve cannot be determined at sight. The bottom line shows their frequencies automatically recorded every 10 seconds. Note the accidental similarity of this curve to the EEG record of alpha rhythms in Figure 3 (b).

ried the physicist in his attempt to understand the composition of our atomic being, into vistas of ever increasing enchantment but describable only in the convention of mathematical language. Today, as we travel from one fresh vista to another, the propriety of the language we use, the convention we adopt, becomes increasingly important. Arithmetic is an adequate language for describing the height and time of the tides, but if we want to predict their rise and fall we have to use a different language, an algebra, with its special notation and theorems. In similar fashion, the electrical waves and tides in the brain can be described adequately by counting, by arithmetic; but there are many unknown quantities when we come to the more ambitious purposes of understanding and predicting brain behaviour—many *x*'s and *y*'s; so it will have to have its algebra. The word is forbidding to some people; but, after all, it means no more than "the putting together of broken pieces."

EEG records may be considered, then, as the bits and pieces of a mirror for the brain, itself *speculum speculorum*. They must be carefully sorted before even trying to fit them together with bits from other sources. Their information comes as a conventional message, coded. You may crack the code, but that does not imply that the information will necessarily be of high significance. Supposing, for instance, you pick up a coded message which you think may be about a momentous political secret. In the first stage of decoding it you might ascertain that the order of frequency of the letters was *ETAONI*. This does not sound very useful information; but reference to the letter-frequency tables would assure you at least that it was a message in English and possibly intelligible. Likewise, we watch the frequencies as well as the amplitude and origin of the brain rhythms, knowing that many earnest seekers for the truth have spent lifetimes trying to decipher what they thought were real messages, only to find that their horoscopes and alembics contained gibberish. The scientist is used to such hazards of research; it is only the ignorant and superstitious who regard him, or think he regards himself, as a magician or priest who is right about everything all the time.

Brain research has just about reached the stage where the letter frequencies of the code indicate intelligibility and their grouping significance. But there is this complication. The ordinary coded message is a sequence in time; events in the brain are not a single sequence in time—they occur in three-dimensional space, in that one bit of space which is more crowded with events than any other we can conceive. We may tap a greater number of sectors of the brain and set more pens scribbling; but the effect of this will only be to multiply the number of code signals, to the increasing embarrassment of the observer, unless the order and inter-relation of the signals can be clarified and emphasised. Redundancy is already a serious problem of the laboratory.

The function of a nervous system is to receive, correlate, store and generate many signals. A human brain is a mechanism not only far more intricate than any other but one that has a long individual history. To study such a problem in terms of frequency and amplitude as a limited function of time—in wavy lines—is at the best over-simplification. And the redundancy is indeed enormous. Information at the rate of about 3,600 amplitudes per minute may be coming through each of the eight channels during the average recording period of 20 minutes; so the total information in a routine record may be represented by more than half a million numbers; yet the usual description of a record consists only of a few sentences. Only rarely does an observer use more than one-hundredth of one per cent of the available information.

"What's in a brain that ink may character . . . ?"

For combining greater clarity with greater economy, many elaborations of methods have been adopted in clinic and laboratory. They still do not overcome the fundamental embarrassment of redundancy and the error of over-simplification, both due to the limitations of a time scale. A promising alternative is a machine that draws a snapshot map instead of a long history, projecting the electrical data visually on a spatial co-ordinate system which can be laid out so as to represent a simple map or model of the head. This moving panorama of the brain rhythms does approximate to Sherrington's "enchanted loom where millions of flashing shuttles weave a dissolving pattern, always a meaningful pattern though never an abiding one." (Figure 5.)

We have called the apparatus which achieves this sort of effect at the Burden Institute a toposcope, by reason of its display of topographic detail. The equipment was developed by Harold Shipton, whose imaginative engineering transformed the early models from entertainment to education. Two of its 24 channels are for monitoring the stimuli; the others, instead of being connected with pens, lead the electrical activity of the brain tapped by the electrodes for display on the screens of small cathode-ray tubes. So instead of wavy lines on a moving paper, the observer sees, to quote Sherrington again, "a sparkling field of rhythmic flashing points with trains of travelling sparks hurrying hither and thither." Assembled in the display console, 22 of the tubes give a kind of Mercator's projection of the brain. Frequency, phase and time relations of the rhythms are shown in what at first appears to be a completely bewildering variety of patterns in each tube and in their ensemble. Then, as the practised eye gains familiarity with the scene, many details of brain activity are seen for the first time. A conventional pen machine is simultaneously at the disposal of the observer, synchronised so that, by turning a switch, a written record of the activity seen in any five of the tubes can be made. Another attachment is a camera with which at the same time permanent snapshot records of the display can be obtained. (Figure 6.)

Thus, from Berger's crude galvanometer to this elaborate apparatus requiring a whole room of its own, electroencephalography has progressed from a technique to a science. Its clinical benefits, by-products of free research, are acknowledged; they can be gauged by the vast multiplication of EEG laboratories. From Berger's lone clinic have sprung several hundred EEG centres—more than 50 in England alone. Literally millions of yards of paper have been covered with frantic scribblings. In every civilized country there is a special learned society devoted to the discussion of the records and to disputation on technique and theory. These societies are banded together in an International Federation, which publishes a quarterly Journal and organises international congresses.

For a science born, as it were, bastard and neglected in infancy, this is a long way to have travelled in its first quarter of a century. If it is to provide the mirror which the brain requires to see itself steadily and whole, there is still a long road ahead. The following chapters give the prospect as seen from the present milestone, assuming that such studies are allowed to continue. Looking back, we realise that the present scale of work as compared with previous physiological research is elaborate and expensive. But our annual cost of conducting planned investigations of a fundamental nature into man's supreme faculties is less than half that of one medium tank, and the money spent on brain research in all England is barely one-tenth of one per cent of the cost of the national mental health services.

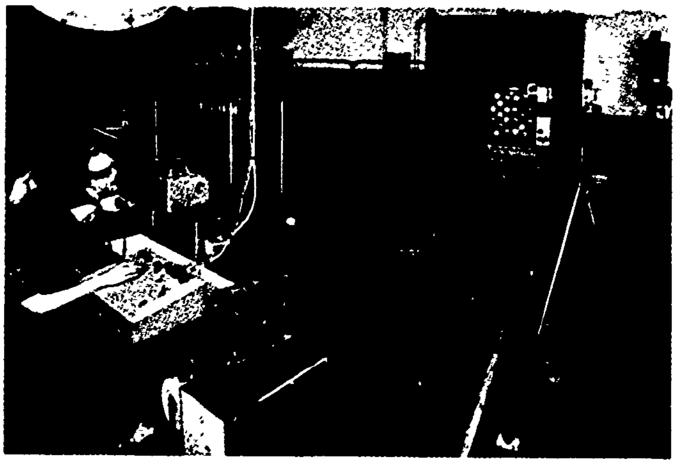


Figure 5. ". . . a moving panorama of the brain rhythms." The Toposcope Laboratory. The subject's couch and triggered stroboscope (flicker) reflector at extreme left beyond desk of 6-channel pen recorder with remote control panel. The 22-channel toposcope amplifier is in the background, the display panel at right centre, camera and projector at extreme right.

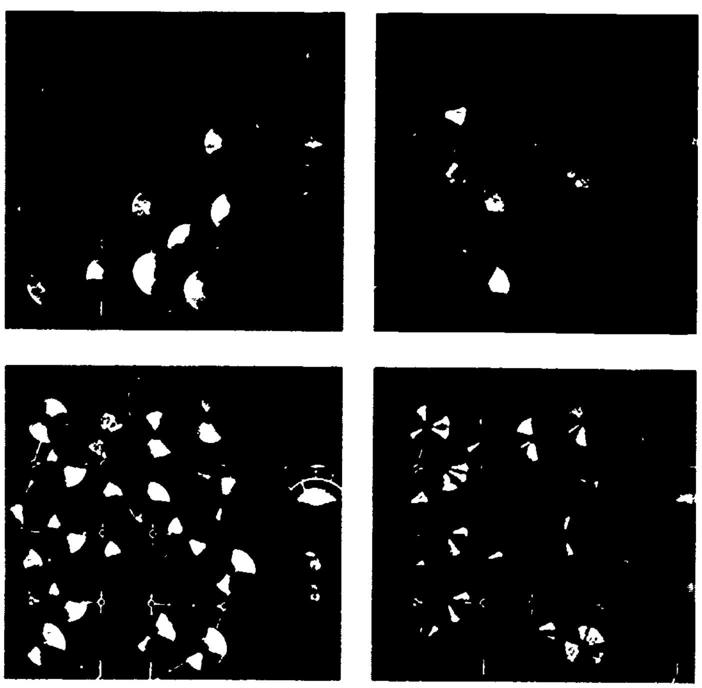
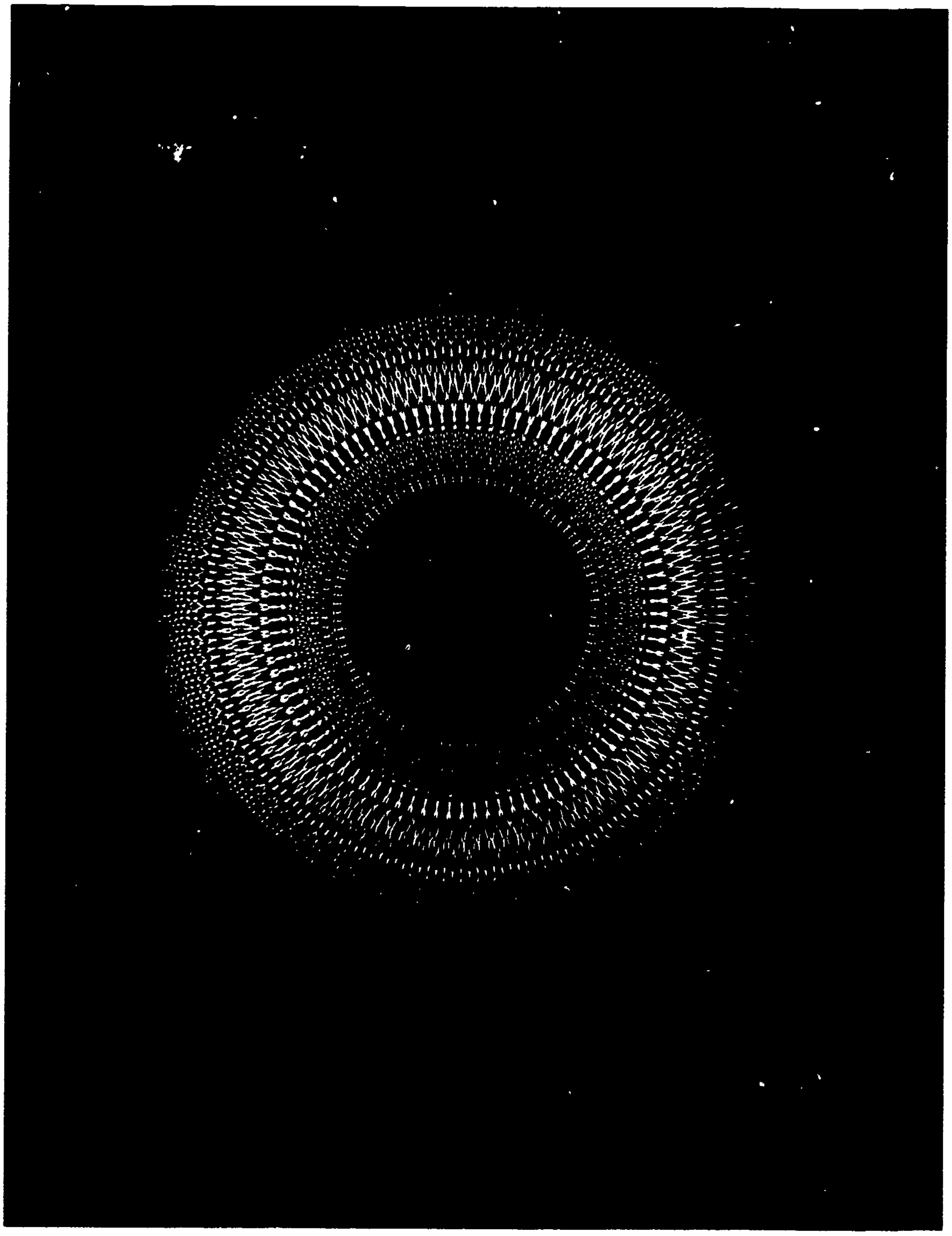


Figure 6. ". . . always a meaningful pattern though never an abiding one." Snapshots of the "sparkling field of rhythmic flashing points." Each of the tube screens, which form a chart of the head seen from above with nose at top, shows by the flashing sectors of its disc the activity of the corresponding area of the brain. (*Top left*) Resting alpha rhythms. (*Top right*) Theta rhythms in anger. (*Bottom left*) Wide response to double flashes of light. (*Bottom right*) Spread of response to triple flashes.



Antenna pattern simulation.

[illegible] is full of concepts [illegible] pictures. [illegible] scientists, [illegible] mathematical [illegible].

18 Scientific Imagination

Richard P. Feynman, Robert B. Leighton, and Matthew Sands

Excerpt from *The Feynman Lectures on Physics,* Volume II, 1964.

I have asked you to imagine these electric and magnetic fields. What do you do? Do you know how? How do *I* imagine the electric and magnetic field? What do *I* actually see? What are the demands of scientific imagination? Is it any different from trying to imagine that the room is full of invisible angels? No, it is not like imagining invisible angels. It requires a much higher degree of imagination to understand the electromagnetic field than to understand invisible angels. Why? Because to make invisible angels understandable, all I have to do is to alter their properties *a little bit*—I make them slightly visible, and then I can see the shapes of their wings, and bodies, and halos. Once I succeed in imagining a visible angel, the abstraction required—which is to take almost invisible angels and imagine them completely invisible—is relatively easy. So you say, "Professor, please give me an approximate description of the electromagnetic waves, even though it may be slightly inaccurate, so that I too can see them as well as I can see almost invisible angels. Then I will modify the picture to the necessary abstraction."

I'm sorry I can't do that for you. I don't know how. I have no picture of this electromagnetic field that is in any sense accurate. I have known about the electromagnetic field a long time—I was in the same position 25 years ago that you are now, and I have had 25 years more of experience thinking about these wiggling waves. When I start describing the magnetic field moving through space, I speak of the ***E***- and ***B*** fields and wave my arms and you may imagine that I can see them.

I'll tell you what I see. I see some kind of vague shadowy, wiggling lines—here and there is an E and B written on them somehow, and perhaps some of the lines have arrows on them—an arrow here or there which disappears when I look too closely at it. When I talk about the fields swishing through space, I have a terrible confusion between the symbols I use to describe the objects and the objects themselves. I cannot really make a picture that is even nearly like the true waves. So if you have some difficulty in making such a picture, you should not be worried that your difficulty is unusual.

Our science makes terrific demands on the imagination. The degree of imagination that is required is much more extreme than that required for some of the ancient ideas. The modern ideas are much harder to imagine. We use a lot of tools, though. We use mathematical equations and rules, and make a lot of pictures. What I realize now is that when I talk about the electromagnetic field in space, I see some kind of a superposition of all of the diagrams which I've ever seen drawn about them. I don't see little bundles of field lines running about because it worries me that if I ran at a different speed the bundles would disappear. I don't even always see the electric and magnetic fields because sometimes I think I should have made a picture with the vector potential and the scalar potential, for those were perhaps the more physically significant things that were wiggling.

Perhaps the only hope, you say, is to take a mathematical view. Now what is a mathematical view? From a mathematical view, there is an electric field vector and a magnetic field vector at every point in space; that is, there are six numbers associated with every point. Can you imagine six numbers associated with each point in space? That's too hard. Can you imagine even *one* number associated with every point? I cannot! I can imagine such a thing as the temperature at every point in space. That seems to be understandable. There is a hotness and coldness that varies from place to place. But I honestly do not understand the idea of a *number* at every point.

So perhaps we should put the question: Can we represent the electric field by something more like a temperature, say like the displacement of a piece of jello? Suppose that we were to begin by imagining that the world was filled with thin jello and that the fields represented some distortion—say a stretching or twisting—of the jello. Then we could visualize the field. After we "see" what it is like we could abstract the jello away. For many years that's what people tried to do. Maxwell, Ampere, Faraday, and others tried to understand electromagnetism this way. (Sometimes they called the abstract jello "ether.") But it turned out that the attempt to imagine the electromagnetic field in that way was really standing in the way of progress. We are unfortunately limited to abstractions, to using instruments to detect the field, to using mathematical symbols to describe the field, etc. But nevertheless, in some sense the fields are real, because after we are all finished fiddling around with mathematical equations—with or without making pictures and drawings or trying to visualize the thing—we can still make the instruments detect the signals from Mariner II and find out about galaxies a billion miles away, and so on.

The whole question of imagination in science is often misunderstood by people in other disciplines. They try to test our imagination in the following way. They say, "Here is a picture of some people in a situation. What do you imagine will happen next?" When we say, "I can't imagine," they may think we have a weak imagination. They overlook the fact that whatever we are *allowed* to imagine in science must be *consistent with everything else we know:* that the electric fields and the waves we talk about are not just some happy thoughts which we are free to make as we wish, but ideas which must be consistent with all the laws of physics we know. We can't allow ourselves to seriously imagine things which are obviously in contradiction to the known laws of nature. And so our kind of imagination is quite a difficult game. One has to have the imagination to think of something that has never been seen before, never been heard of before. At the same time the thoughts are restricted in a strait jacket, so to speak, limited by the conditions that come from our knowledge of the way nature really is. The problem of creating something which is new, but which is consistent with everything which has been seen before, is one of extreme difficulty.

While I'm on this subject I want to talk about whether it will ever be possible to imagine *beauty* that we can't *see*. It is an interesting question. When we look at a rainbow, it looks beautiful to us. Everybody says, "Ooh, a rainbow." (You see how scientific I am. I am afraid to say something is beautiful unless I have an experimental way of defining it.) But how would we describe a rainbow if we were blind? We *are* blind when we measure the infrared reflection coefficient of sodium chloride, or when we talk about the frequency of the waves that are coming from some galaxy that we can't see—we make a diagram, we make a plot. For instance, for the rainbow, such a plot would be the intensity of radiation vs. wavelength measured with a spectrophotometer for each direction in the sky. Generally, such measurements would give a curve that was rather flat. Then some day, someone would discover that for certain conditions of the weather, and at certain angles in the sky, the spectrum of intensity as a function of wavelength would behave strangely; it would have a bump. As the angle of the instrument was varied only a little bit, the maximum of the bump would move from one wavelength to another. Then one day the physical review of the blind men might publish a technical article with the title "The Intensity of Radiation as a Function of Angle under Certain Conditions of the Weather." In this article there might appear a graph such as the one in Fig. 20-5. The author would perhaps remark that at the larger angles there was more radiation at long wavelengths, whereas for the smaller angles the maximum in the radiation came at shorter wavelengths. (From our point of view, we would say that the light at 40° is predominantly green and the light at 42° is predominantly red.)

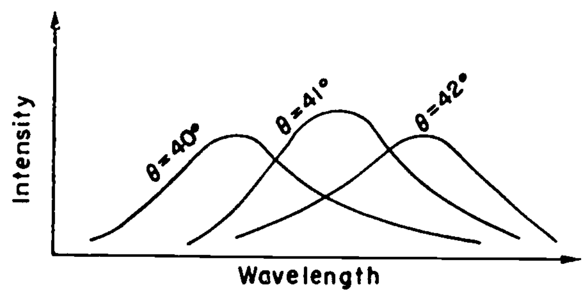


Fig. 20–5. The intensity of electromagnetic waves as a function of wavelength for three angles (measured from the direction opposite the sun), observed only with certain meteorological conditions.

Now do we find the graph of Fig. 20–5 beautiful? It contains much more detail than we apprehend when we look at a rainbow, because our eyes cannot see the exact details in the shape of a spectrum. The eye, however, finds the rainbow beautiful. Do we have enough imagination to see in the spectral curves the same beauty we see when we look directly at the rainbow? I don't know.

But suppose I have a graph of the reflection coefficient of a sodium chloride crystal as a function of wavelength in the infrared, and also as a function of angle. I would have a representation of how it would look to my eyes if they could see in the infrared—perhaps some glowing, shiny "green," mixed with reflections from the surface in a "metallic red." That would be a beautiful thing, but I don't know whether I can ever look at a graph of the reflection coefficient of NaCl measured with some instrument and say that it has the same beauty.

On the other hand, even if we cannot see beauty in particular measured results, we *can* already claim to see a certain beauty in the equations which describe general physical laws. For example, in the wave equation (20.9), there's something nice about the regularity of the appearance of the x, the y, the z, and the t. And this nice symmetry in appearance of the x, y, z, and t suggests to the mind still a greater beauty which has to do with the four dimensions, the possibility that space has four-dimensional symmetry, the possibility of analyzing that and the developments of the special theory of relativity. So there is plenty of intellectual beauty associated with the equations.

NORMAN LEADER ALLEN

Norman Leader Allen, British physicist, was born in 1927 and received his B.Sc. from the University of Birmingham, England, in 1948 and his Ph.D. in 1951. Allen has been a staff member of Massachusetts of Technology and is now a lecturer in the Electrical and Electronic Engineering Department at the University of Leeds. In addition to his book, Threshold Pressure for Arc Discharges, he has written extensively in scientific journals on arc discharges, cosmic rays and plasma physics.

STANLEY SUMNER BALLARD

Stanley S. Ballard, Professor of Physics and chairman of the department at the University of Florida, Gainesville, was born in Los Angeles in 1908. He received his A.B. from Pomona College, and M.A. and Ph.D. from the University of California. He has taught at the University of Hawaii, Tufts University, and has been a research physicist at the Scripps Institution of Oceanography. Ballard is a member of the Physical Science Division, National Research Council, and belongs to several optical societies in this country and Europe, having served as president of the Optical Society of America. His specialities are spectroscopy, optical and infrared instrumentation, and properties of optical materials. Ballard is coauthor of Physical Principles.

ARTHUR C. CLARKE

Arthur C. Clarke, British scientist and writer, is a Fellow of the Royal Astronomical Society. During World War II he served as technical officer in charge of the first aircraft ground-controlled approach project. He has won the Kalinga Prize, given by UNESCO for the popularization of science. The feasibility of many of the current space developments was perceived and outlined by Clarke in the 1930's. His science fiction novels include Childhoods End and The City and the Stars.

ALBERT EINSTEIN

Albert Einstein, considered to be the most creative physical scientist since Newton, was nevertheless a humble and sometimes rather shy man. He was born in Ulm, Germany, in 1879. He seemed to learn so slowly that his parents feared that he might be retarded. After graduating from the Polytechnic Institute in Zurich, he became a junior official at the Patent Office at Berne. At the age of twenty-six, and quite unknown, he published three revolutionary papers in theoretical physics in 1905. The first paper extended Max Planck's ideas of quantization of energy, and established the quantum theory of radiation. For this work he received the Nobel Prize for 1929. The second paper gave a mathematical theory of Brownian motion, yielding a calculation of the size of a molecule. His third paper founded the special theory of relativity. Einstein's later work centered on the general theory of relativity. His work had a profound influence not only on physics, but also on philosophy. An eloquent and widely beloved man, Einstein took an active part in liberal and anti-war movements. Fleeing from Nazi Germany, he settled in the United States in 1933 at the Institute for Advanced Study in Princeton. He died in 1955.

RICHARD PHILLIPS FEYNMAN

Richard Feynman was born in New York in 1918, and graduated from the Massachusetts Institute of Technology in 1939. He received his doctorate in theoretical physics from Princeton in 1942, and worked at Los Alamos during the Second World War. From 1945 to 1951 he taught at Cornell, and since 1951 has been Tolman Professor of Physics at the California Institute of Technology. Professor Feynman received the Albert Einstein Award in 1954, and in 1965 was named a Foreign Member of the Royal Society. In 1966 he was awarded the Nobel Prize in Physics, which he shared with Shinichero Tomonaga and Julian Schwinger, for work in quantum field theory.

LEOPOLD INFELD

Leopold Infeld, a co-worker with Albert Einstein in general relativity theory, was born in 1898 in Poland. After studying at the Cracow and Berlin Universities, he became a Rockefeller Fellow at Cambridge where he worked with Max Born in electromagnetic theory, and then a member of the Institute for Advanced Study at Princeton. For eleven years he was Professor of Applied Mathematics at the University of Toronto. He then returned to Poland and became Professor of Physics at the University of Warsaw and until his death on 16 January 1968 he was director of the Theoretical Physics Institute at the university. A member of the presidium of the Polish Academy of Science, Infeld conducted research in theoretical physics, especially relativity and quantum theories. Infeld was the author of The New Field Theory, The World in Modern Science, Quest, Albert Einstein, and with Einstein The Evolution of Physics.

THOMAS JEFFERSON

Thomas Jefferson, third President of the United States, was born in 1743 at Shadwell in Goochland County, Virginia. He studied Greek, Latin, and mathematics at the College of William and Mary for two years, and later became a lawyer. From 1768 to 1775 Jefferson was a member of the Virginia House of Burgesses. In 1775 he was elected to the Second Continental Congress, and in 1776 he drafted the Declaration of Independence. Jefferson felt a conflicting devotion to the tranquil pursuits of science and public service. His interests ranged over such fields as agriculture, meteorology, paleontology, ethnology, botany, and medicine. He believed in the freedom of the scientific mind and the importance of basing conclusions on observations and experiment. Jefferson demanded utility of science, hence his numerous inventions and interest in improvements and simplifications of agricultural tools and techniques, and in balloons, dry docks, submarines, even the furniture in his home (swivel chairs and music stands). Because of his prominence as a public figure, he was influential in increasing and improving science education in America. He died on July 4, 1826, the fiftieth anniversary of the Declaration of Independence.

MATTHEW JOSEPHSON

Matthew Josephson, prolific writer and magazine editor, was born in Brooklyn in 1899. He received his B.A. from Columbia University in 1920. Josephson was successively editor of the Broom, Transition, and The New Republic, which he left in 1932. In 1948 he was elected to the National Institute of Arts and Letters and also was a traveling Guggenheim fellow for creative literature. He is the author of Zola and His Time, The Robber Barons, and Portrait of the Artist As American.

ROBERT B. LEIGHTON

Robert B. Leighton, born in Detroit, Michigan in 1919, was first a student and then a faculty member at California Institute of Technology. He is a member of the International Astronomical Union, the National Academy of Science and the American Physics Society. Professor Leighton's work deals with the theory of solids, cosmic rays, high energy physics, and solar physics.

DAVID KEITH CHALMERS MACDONALD

David Keith Chalmers MacDonald was born in Glasgow, Scotland, in 1920 and received his M.A. in mathematics and natural philosophy from Edinburgh University in 1941. After serving with the Royal Mechanical and Electrical Engineers during World War II, he received his Ph.D. in 1946 from Edinburgh. Then he attended Oxford as a research fellow and received a Ph.D. in 1949. In 1951 Dr. MacDonald went to Canada and started a low temperature physics research laboratory for the National Research Council. MacDonald was appointed to the physics department at Ottawa University in 1955 and elected Fellow of the Royal Society of London in 1960. Aside from numerous articles in scientific journals, he was the author of Near Zero: An Introduction to Low Temperature Physics and Faraday, Maxwell, and Kelvin. MacDonald died in 1963.

JAMES CLERK MAXWELL

See J. R. Newman's articles in Readers 3 and 4.

ALBERT ABRAHAM MICHELSON

Precision measurement in experimental physics was the lifelong passion of A. A. Michelson (1852-1931), who became in 1907 the first American to win a Nobel Prize in one of the sciences. Born in Prussia but raised in California and Nevada, Michelson attended the U. S. Naval Academy and was teaching there in 1879 when he first improved the methods of measuring the velocity of light on earth. After a post-graduate education in Europe under several great professors of optics, Michelson returned to the United States where he taught physics at the college that became Case Institute of Technology, then at Clark University, and at the University of Chicago. While in Europe he invented the famous instrument called the Michelson interferometer and while in Cleveland at Case in 1887, he and E. W. Morley improved this device in an effort to measure the absolute velocity of the Earth as it hurtles through space. The failure of the Michelson-Morley aether-drift experiment was an important part of the background leading to Albert Einstein's first work on the theory of relativity. Although Michelson remained a creative experimentalist in physical optics, meteorology, astrophysics and spectroscopy throughout his life, he died still believing exclusively in the wave model of the nature of light and in his "beloved aether." His figures for the velocity of light, refined still further just before his death, remain the accepted value of one of the few "absolute" constants in physics today.

JAMES ROY NEWMAN

James R. Newman, lawyer and mathematician, was born in New York City in 1907. He received his A.B. from the College of the City of New York and LL.B. from Columbia. Admitted to the New York bar in 1929, he practiced there for twelve years. During World War II he served as chief intelligence officer, U. S. Embassy, London, and in 1945 as special assistant to the Senate Committee on Atomic Energy. From 1956-57 he was senior editor of The New Republic, and since 1948 had been a member of the board of editors for Scientific American where he was responsible for the book review section. At the same time he was a visiting lecturer at the Yale Law School. J. R. Newman is the author of What is Science?, Science and Sensibility, and editor of Common Sense of the Exact Sciences, The World of Mathematics, and the Harper Encyclopedia of Science. He died in 1966.

MATTHEW SANDS

Matthew Sands was born in Oxford, Massachusetts, in 1919 He attended Clark College, Rice Institute, and Massachusetts Institute of Technology. During World War II he worked at the Los Alamos Scientific Laboratory. He was Professor of Physics at the California Institute of Technology before joining the linear accelerator group at Stanford University. Professor Sands specializes in electronic instrumentation for nuclear physics, cosmic rays, and high-energy physics. He served as chairman of the Commission on College Physics.

WILLIAM ASAHEL SHURCLIFF

Born in Boston in 1909, William A. Shurcliff was educated at Harvard, receiving his Ph.D. in physics in 1934. During the war he served as technical aide to the Office of Scientific Research and Development, National Defense Research Committee, and Manhattan project. Then he was with the Polaroid Corporation as senior scientist and project leader. He is now a Research Fellow at the Electron Accelerator at Harvard. Shurcliff is the author of Polarized Light: Production and Use and Bombs at Bikini. His technical interests include emission spectroscopy, absorption spectrophotometry, atomic energy, gamma radiation dosimeters, microscope design, and color vision.

JAMES ALFRED VAN ALLEN

James Alfred Van Allen, discoverer of the "Van Allen radiation belt," was born at Mt. Pleasant, Iowa, in 1914. After his undergraduate work at Iowa Wesleyan College, he received his M.S. and in 1939 his Ph.D. from the State University of Iowa, where he is now a Professor of Physics and Astronomy. He has been a Carnegie research fellow at the Carnegie Institution, Guggenheim fellow, and a research associate at Princeton, and is the recipient of numerous honorary doctorates. For his distinguished work in nuclear physics, cosmic rays and space probes, he has been awarded the Hickman Medal from the American Rocket Society, the Distinguished Civilian Service Medal of the U.S. Army, and the Hill Award of the Institute of Aerospace Science.

EDGAR VILLCHUR

Edgar Villchur is President and Director of Research of the Foundation for Hearing Aid Research in Woodstock, New York. He was born in New York City in 1917 and received a M.S.Ed. from the City College of New York. He has taught at New York University, and was President and Chief Designer of Acoustic Research, Inc., a manufacturing company in the high fidelity field.

WILLIAM GREY WALTER

William Grey Walter was born in 1911 and received his M.A. and Sc.D. (1947) from Cambridge University. He was a Rockefeller Fellow at the Maudsley Hospital in England. W. Grey Walter is a pioneer in the use of electroencephalography for translating the minute electrical currents of the human brain into physical patterns which may be studied for the information they give us on brain processes. Walter is the author of The Living Brain, Further Outlook, The Curve of the Snowflake and articles to various scientific journals.

THOMAS YOUNG

Thomas Young, the versatile English physician and physicist, was born in Milverton, Somersetshire, England, in 1773. He became Professor of Natural Philosophy at the Royal Institution in 1801 and was foreign secretary of the Royal Society from 1802 to 1829. His most important contribution to science was his discovery of the interference of light and the measurements of the wavelengths of various colors. Others included the definition of energy, study of sound and elasticity (Young's Modulus), discovery of astigmatism (in himself), explanation of accommodation, and initiation of a color vision theory later developed by Helmholtz. His interest in linguistics led him to participate in the deciphering of Egyptian hieroglyphics. Thomas Young died in 1829.

1. LETTER FROM THOMAS JEFFERSON, JUNE, 1799, from Scripta Mathematica, Volume 1, 1932, pages 88-90. Permission to publish by Teachers College Library, Columbia University. Also in Science and the Common Understanding, by J. Robert Oppenheimer, 1953. Reprinted by permission of Simon and Schuster, Inc.

2. ON THE METHOD OF THEORETICAL PHYSICS, by Albert Einstein, from Essays in Science, pages 12-21, Philosophical Library, New York, Copyright (c) 1934. Reprinted with permission.

3. EXPERIMENTS AND CALCULATIONS RELATIVE TO PHYSICAL OPTICS, by Thomas Young, from Miscellaneous Works of Thomas Young, Volume 1, edited by George Peacock, John Murray, London, 1855.

4. VELOCITY OF LIGHT, by A. A. Michelson, from Studies in Optics, The University of Chicago Press, 1927. Reprinted with permission of the University of Chicago.

5. POPULAR APPLICATIONS OF POLARIZED LIGHT, by William A. Shurcliff and Stanley S. Ballard, from Polarized Light (Momentum #7), Copyright 1964, D. Van Nostrand Company, Inc., Princeton, New Jersey. Reprinted with permission.

6. ACTION AT A DISTANCE, by James Clerk Maxwell, from The Scientific Papers of James Clerk Maxwell, Volume I, edited by D. W. Niven, Cambridge University Press, Cambridge, England, 1890. Reprinted with permission.

7. THE ELECTRONIC REVOLUTION, by Arthur C. Clarke, from Voices From The Sky, pages 204-210, Harper and Row, Publishers, New York, 1965. "The Electronic Revolution" copyright (c) 1962 by The New York Times Company. Reprinted by permission of the author and his agents: Scott Meredith Literary Agency, Inc., 580 Fifth Avenue, New York 10036.

8. THE INVENTION OF THE ELECTRIC LIGHT, by Matthew Josephson. Reprinted with permission. Copyright (c) 1959 by Scientific American, Inc. All rights reserved.

9. HIGH FIDELITY, by Edgar Villchur, from Reproduction of Sound, Chapters 2 and 3, AR Library Volume 2; Copyright 1962 by Acoustic Research, Inc., Cambridge, Massachusetts. Copyright 1965 by Dover Publications, Inc., New York, New York. Reprinted with permission of the publisher.

10. THE FUTURE OF DIRECT CURRENT POWER TRANSMISSION, by N. L. Allen, from Endeavour, Volume XXVI, No. 97, January 1967., Imperial Chemical Industries Limited, London. Reprinted with permission.

11. JAMES CLLRK MAXWELL, by James R. Newman, Part II. Reprinted with permission. Copyright 1955 by Scientific American, Inc. All rights reserved.

12. MAXWELL'S LETTERS: A COLLECTION

13. ON THE INDUCTION OF ELECTRIC CURRENTS, by James Clerk Maxwell, from A Treatise on Electricity and Magnetism, Volume 2, 1873, The Clarendon Press, Oxford. Reprinted with permission.

14. THE RELATIONSHIP OF ELECTRICITY AND MAGNETISM, by D. K. C. MacDonald, from Faraday, Maxwell, and Kelvin. Copyright 1964 by Educational Services Incorporated. Reprinted by permission of Doubleday & Company, Inc. (Science Study Series.)

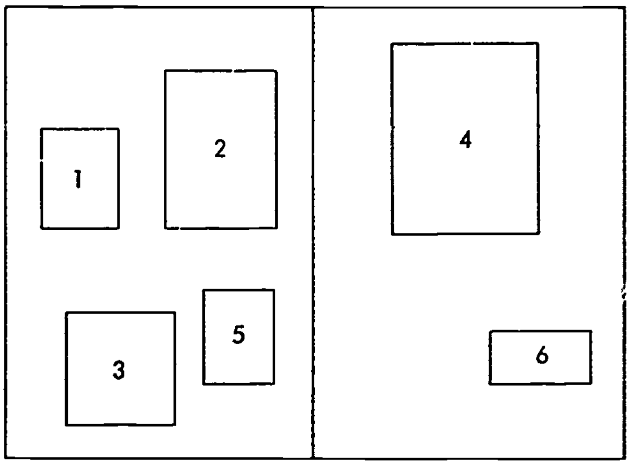
15. THE ELECTROMAGNETIC FIELD, by Albert Einstein and Leopold Infeld, from The Evolution of Physics, Simon and Schuster, New York, 1961. Reprinted with permission.

16. RADIATION BELTS AROUND THE EARTH, by James A. Van Allen. Reprinted with permission. Copyright © 1959 by Scientific American, Inc. All rights reserved. Available separately at 20¢ each as Offprint No. 248 from W. H. Freeman and Company, 660 Market Street, San Francisco, California.

17. A MIRROR FOR THE BRAIN, by W. Grey Walter, from The Living Brain, W. W. Norton and Company, Inc., New York, 1963. Reprinted with permission.

18. SCIENTIFIC IMAGINATION, by Richard P. Feynman, Robert B. Leighton, and Matthew Sands, from The Feynman Lectures on Physics, Volume II, Addison-Wesley, Reading, Massachusetts, 1964. Reprinted with permission.

Cover: Current, 1964, by Bridget Riley. Emulsion on composition board, 58 3/8 x 58 7/8". Courtesy of The Museum of Modern Art, New York City.



(1) Glen J. Pearcy.
(2) Jeune fille au corsage rouge lisant. Jean Baptiste Camille Corot. Painting. Collection Bührle, Zurich.
(3) HPP staff photo.
(4) Femme lisant. Georges Seurat. Conté crayon drawing. Collection C. F. Stoop, London.
(5) Portrait of Pierre Reverdy. Pablo Picasso. Etching. Museum of Modern Art, N.Y.C.
(6) Lecture au lit. Paul Klee. Drawing. Klee Foundation.

P. 28 Courtesy of The Metropolitan Museum of Art, N.Y.C.

P. 48 c Lincoln Center for the Performing Arts, Inc., 1967—photograph by Morris Warman.

P. 82 Courtesy of Mr. Harold M. Waage, Palmer Physical Laboratory, Princeton University.

P. 90 U.S. Patent Office Picture Collection.

P. 102 International Business Machines Corp. Photo by Ezra Stoller.

P. 122 Tennessee Valley Authority.

P. 156 American Institute of Physics.

P. 212 Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization, Australia.

P. 238 Conductron Corporation, Ann Arbor, Michigan.

This Reader is the property of Harvard Project Physics. Upon request it is to be returned to Pierce Hall, 29 Oxford Street, Cambridge, Massachusetts 02138.